What lies beneath

A focus on specific biological targets rather than constellations of symptoms heralds a more scientific approach to the treatment of neuropsychiatric disorders.

It is difficult to decide which is the more dispiriting statement: that one person in four suffers from a mental-health problem at some point in their lives, or that the sorry state of care for the millions of people affected is often discussed only in the immediate aftermath of the actions of a single mentally ill individual. One can quibble about statistics and diagnostic standards, but the bottom line is that neuropsychiatric disorders account for one of the greatest burdens of disease in the developed world, yet patients are not receiving the help they need.

Part of the problem is that, for many people, the available therapies simply do not work, and that situation is unlikely to improve any time soon. By the early 1990s, the pharmaceutical industry had discovered — mostly through luck — a handful of drug classes that today account for most mental-health prescriptions. Then the pipeline ran dry. On close inspection, it was far from clear how the available drugs worked. Our understanding of mental-health disorders, the firms realized, is insufficient to inform drug development.

With industry pulling out, the burden of developing therapies for mental disorders is increasingly falling on academic researchers. And with its strapped budget, it comes as no surprise that the US National Institute of Mental Health (NIMH) in Bethesda, Maryland, wants to make quick and decisive changes to the way its money is spent.

For several years, the NIMH has been trying to forge a different approach, and late last month institute director Thomas Insel announced that the agency will no longer fund clinical trials that do not attempt to determine a drug or psychotherapy's mechanism of action (see page 288). Without understanding how the brain works, he has long maintained, we cannot hope to know how a therapy works.

According to the NIMH's new grant framework, clinical trials must be designed to yield scientific knowledge as well as to determine whether a therapy works. This will allow researchers to decide quickly whether to proceed with a trial, modify it or end it before too much money is spent on recruiting and testing patients. To receive funding, scientists will have to show how they intend to test whether a therapy targets a particular brain circuit, for instance, or what mechanism a therapy uses to alter a person's behaviour.

The move will certainly ruffle feathers. Insel notes that more than half of the trials that the NIMH currently supports would not receive funding under the new requirements, at least not without modifications. For example, a trial that focuses on changes in attention span as a means of testing a behavioural-intervention therapy in children with a broadly defined disorder such as attention-deficit hyperactivity disorder would not be funded unless researchers could present a controlled way to study how the therapy takes effect. The NIMH would rather see trials that aim to recruit people with a common trait — say, hallucinations — regardless of their specific psychiatric diagnosis, treat them with a drug that acts on a specific brain receptor, and measure changes in brain activity.

The change applies to trials of drugs and behavioural therapies alike: the key is that they are set up in such a way that a single variable can be changed and a single effect seen. An example would be deep brain stimulation (DBS), which has proved an effective treatment for disorders ranging from Parkinson's disease to depression, but whose underlying mechanism is largely unknown. When a major trial of DBS for depression failed last year, scientists were left not knowing whether they had sited the devices wrongly or recruited the wrong group of patients. It is becoming increasingly apparent that the optimal placement of and frequency used by such devices to stimulate neurons may differ from person to person (see page 290).

> ***"Our understanding of mental-health disorders is insufficient to inform drug development."***

Critics will argue that the NIMH has exchanged a difficult problem — treating mental illness — for an even more challenging one, understanding the brain. But the institute's new direction on trials may also aid an effort to free research on mental-health disorders from the limits of existing diagnostic categories — an approach for which *Nature* has previously expressed support (see *Nature* **496,** 397–398; 2013).

A project called the Research Domain Criteria, which the NIMH is developing, places patients on a matrix on the basis of both their symptoms and biomarkers such as brain activity. According to the NIMH, more and more grant applicants are using these criteria, although no trials have yet been completed. If this work throws up new targets for possible intervention, industry may once again see psychiatric drugs as a market worth investing in. It is a gamble — and a variety of approaches are still needed — but the potential pay-off is worth it. ■

How to get ahead

The success of the \$1,000 genome programme offers lessons for fostering innovation.

Genome technologists are on the verge of fulfilling a goal that was once considered so far off as to be almost fictional: the sequencing of human genomes for US\$1,000 apiece. Eleven years ago, when genome sequencing cost tens of millions of dollars, scientists from the US National Human Genome Research Institute (NHGRI) in Bethesda, Maryland, said that, should it be met, the \$1,000 goal "would revolutionize biomedical research and clinical practice".

The price of sequencing has indeed dropped — and this development is changing science and beginning to change medicine. In retrospect, it may seem that it was inevitable that the cost would come down from what it was at the close of the Human Genome Project

in 2003. Entrepreneurs and scientists were already working on technologies to replace Sanger sequencing, the workhorse of the genome project. But these entrepreneurs and other genome technologists give at least some of the credit for the precipitous price drop to the NHGRI. In 2003 it launched the $230-million Advanced Sequencing Technology programme, to support research to improve genome technology in academia and industry (see page 294). The programme will give out its final grants this year, but it has provided some key lessons on how to set up similar sponsored innovation efforts.

Set a clear goal. The NHGRI programme — often referred to as the $1,000 genome programme — shares this feature with other successful technological endeavours, including the British government's eighteenth-century Longitude Prize for a practical method to determine a ship's position; NASA's Apollo Moon shot in the 1960s; and even the US Defense Advanced Research Projects Agency's contest to sponsor the creation of self-driving cars, first held in 2004. Stating a tangible goal galvanizes the field, harnessing researchers' collective brainpower to achieve a common aim.

Set the bar high, but not too high. The Archon Genomics X Prize was cancelled last year after only two teams signed up to take on its challenge: the sequencing of 100 complete human genomes in 30 days at unparalleled accuracy and low cost. That test was widely seen as too difficult, dissuading researchers from trying. The NHGRI, by contrast, seems to have set a more reasonable goal; before setting $1,000 as its benchmark, it undertook a broad consultation of scientists in the field.

Spur competition. The 97 groups awarded grants by the NHGRI programme so far are made up of academics and companies working on a diverse range of concepts. Recipients have included a who's who of the sequencing industry past and present, as well as a huge array of people working on speculative ideas. Each grant — ranging from hundreds of thousands of dollars to a few million — is small compared with the amount of funding needed to take an idea all the way to market. But the spreading of money between so many competitors keeps the field lively, and keeps market leaders on their toes.

Foster cooperation. The programme convenes all grant recipients every year for a meeting, where all — including those from industry — are expected to describe their progress. Jeffery Schloss, the amiable head of the programme, deserves much of the credit for coaxing grant recipients to share information, and for promoting an atmosphere of collegiality that has moved the field forward.

"Cheaper sequencing is changing science and beginning to change medicine."

Seed a broad range of ideas. The vast majority of the programme's grants — 75% so far — have gone to academics rather than companies. This has nurtured ideas that are too speculative for industry, ensuring that when the market weeds out less successful companies or throws up a monopoly, there is a nursery of new concepts waiting to keep the field from stagnating.

Be flexible. The NHGRI deserves credit for allowing Schloss to operate the programme differently from most of its schemes. For instance, reviewers were allowed to award small chunks of money to parts of an application that seemed promising, rather than cutting off funding for every project that failed to meet a milestone. This allowed the programme to provide more grants than would otherwise have been possible.

Of course, just incorporating these six elements into a programme does not guarantee success. Entrepreneurs and academic scientists have contributed ideas and drive to genomics advances, and venture capitalists have contributed money. And although the $1,000 goal is within striking distance, it has not yet enabled the depth of understanding needed to make full medical or biological use of the knowledge derived from ever more genomes. Attacking that problem is the next challenge of genomics. But in part because of the $1,000 genome programme, biologists are now in a position to address it. ■

Past wisdom

The recent Nature PastCast *series is instructive as well as entertaining.*

Four decades ago, in August 1975, a short paper on antibodies was published in *Nature*. The findings spawned a multibillion-dollar drug industry: the monoclonal antibody drugs Herceptin, used to treat breast cancer, and Humira, used to treat rheumatoid arthritis, were together worth around US$15.4 billion to the pharmaceutical industry in 2012. But the revolution was nearly stillborn: the published technique to make monoclonal antibodies could not be reproduced. The crisis was such, says historian Lara Marks at King's College London, that the authors considered withdrawing their article from *Nature*.

It was years before the investigations paid off and the huge medical potential of monoclonal antibodies started to crystallize. "Nowadays it seems prophetic," says Greg Winter, who made the first monoclonal antibody for use in humans. "But at that point I don't think anyone had realized the importance."

If those who do not learn from the past are truly doomed to repeat the same mistakes, then *Nature* has done its bit over the past few months to help them avoid that fate.

There are many lessons from history in our archive, alongside nuggets of scientific insight and experience. Some of these — and contemporary reflections on them from experts such as Marks — have been presented in the *Nature PastCast*, a 12-part audio series on the history of science that draws to a close this week. Some stories featured in the series are humbling; others simply entertaining. All are relevant.

Are you navigating a tricky international collaboration? So was astronomer Arthur Eddington, who tried to mount an expedition to view a solar eclipse as the First World War was drawing to a close. He was an Englishman — a pacifist Quaker to boot — trying to find evidence to support a German-born physicist's theory (Albert Einstein's general theory of relativity). Or maybe your desk is piled high with a backlog of data needing analysis? In the mid-1980s, climate scientist Jonathan Shanklin had to sift through reams of the stuff before the evidence for the ozone hole popped out at him.

Or are you having trouble avoiding jargon in your latest write-up? In a November 1919 discussion of relativity, the physicist J. J. Thomson is reported to have 'regretted the very complicated form in which Einstein expressed his theory, and hoped it might be possible to put it into a form in which it would be more generally comprehensible and useful'.

Some of the best stories are those that reflect a simple commitment to doing science. In November 1924, palaeontologist Raymond Dart was getting ready to appear as best man at his friend's wedding when a box of fossils arrived at his home in Johannesburg. Having an inkling of its content, he couldn't resist opening the package. In his diary, Dart recounts how the groom himself had to drag him away. "My God, Ray, you've got to finish dressing!"

Wilhelm Röntgen showed similar mettle. According to Otto Glasser's 1934 biography, in the days after he first discovered X-rays, Röntgen ate and slept in his lab "in order to avoid the distracting influence of daily trivialities, and to be able to continue his experiments immediately in case of a sudden inspiration". Once Röntgen had published his X-ray experiments, a journalist asked him what he thought of his new findings. Röntgen gave a reply that every researcher should consider sticking on their fridge. "I did not think; I investigated." ■

➲ NATURE.COM
To comment online, click on Editorials at: go.nature.com/xhunqv

Wanted: Fraud-buster with political antennae

A mild-mannered man's fiery resignation leaves a troubling vacancy at the world's largest office for investigating scientific fraud, says **Colin Macilwain**.

When David Wright leaves the Office of Research Integrity (ORI) at the US Public Health Service later this month, everyone will know just why he is walking out of the door. Wright's resignation letter to the assistant secretary of health, leaked last week, is unusually direct. "I'm offended as an American taxpayer that the federal bureaucracy — at least the part I've labored in — is so profoundly dysfunctional," he tells his former boss, in a note decorated with choice cuts of bureaucratic hubris. In one example, Wright says, he sought an evaluation of the support services available to the ORI, only to be told that "that had been tried a few years ago and the results were so negative that no further evaluations have been conducted".

Wright, a science historian and former research-integrity officer at Michigan State University in East Lansing, signs off with a cheery promise to publish a version of the daily log he kept at the ORI, "to share my experience and observations with my colleagues in government and with members of the regulated research community". That would be you, dear reader.

Anyone who has dealt with Wright professionally will be taken aback to see the guy nailing his colours to the mast in this way. He just isn't the rocking-the-boat type: a more courteous and polite official it would be difficult to meet. There is a profound feeling in circles interested in research misconduct that he is one of the good guys.

"A lot of us are wondering where we go from here," says Mark Frankel, head of the Scientific Responsibility, Human Rights and Law Program at the American Association for the Advancement of Science in Washington DC. As well one might, when the directorship of the world's largest and best-established research-misconduct office — which lacked a permanent director for two years before Wright's appointment in 2012 — has just imploded.

The ORI was established 22 years ago in the wake of the David Baltimore case, in which allegations of fraud (later dismissed) rocked the laboratory of one of America's most eminent biologists. The agency's 25 staff are supposed to educate tens of thousands of researchers on proper research conduct, as well as overseeing investigations into misconduct by researchers funded by the US National Institutes of Health (NIH), the world's largest research agency. But it was built to deal with the handful of cases that it was assumed would arise. According to research published last December by Daniele Fanelli at the University of Edinburgh, UK (D. Fanelli *PLoS Med.* **10,** e1001563; 2013), the proportion of allegations the ORI receives that are investigated and closed has halved over 20 years. Wright's resignation letter blows wide open long-standing doubts about its capacity to deal with a caseload that, the available evidence suggests, should have expanded with the growth of the NIH itself.

THE POSITION MUST BE FILLED PROMPTLY BY SOMEONE RESPECTED BY BOTH ETHICISTS AND HEALTH RESEARCHERS.

The administration of US President Barack Obama needs to get a grip on this before an explosive high-profile case — such as that of Andrew Wakefield and MMR vaccines in the United Kingdom — turns up and the ORI can't cope. If the administration doesn't do this, Congress just might. Senator Chuck Grassley (Republican, Iowa) demanded in February that the ORI explain its lenient treatment of Dong-Pyou Han, a physician at Iowa State University in Ames, who was banned from seeking NIH funding for three years after falsifying data in AIDS vaccine trials that cost the agency US$19 million.

It is not known how much Grassley's strident demand for answers contributed to Wright's departure. But attention from one of the most feared and respected voices on Capitol Hill can only intensify the political hot-house atmosphere that his resignation letter blames for the ORI's troubles.

These troubles go back a long way. The ORI sits under the assistant health secretary, instead of being properly independent like the inspector generals who keep an eye (imperfectly) on scientific fraud at the National Science Foundation (NSF) and other agencies. Its enforcement remit (of fabrication, falsification and plagiarism, or FFP) is too narrow, and its budget too small.

Canada has already shown the way. In 2011, it set up the Tri-Agency Framework: Responsible Conduct of Research. It has a relatively large staff of eight, for a research system one-tenth the size of that of the United States, a remit that reaches far beyond FFP, and it is led by Susan Zimmerman, a tough lawyer unlikely to put up with dodgy academics whimpering about their "creativity".

Wright's resignation comes as a panel chaired by Robert Nerem, a bioengineer at the Georgia Institute of Technology in Atlanta, is in the final stages of the first National Academy of Sciences study on research integrity in 20 years. The report may recommend beefing up the ORI, broadening its remit beyond FFP and aligning its approach with that of the NSF and other agencies. But with the ORI leaderless, the academy report may be even more prone than usual to gathering dust.

I have been an optimist on misconduct: some 18 months ago, I wrote that the global community was starting to get a handle on it. Wright's appointment was progress, and his departure is a setback. His position must be filled promptly by someone respected by both ethicists and health researchers. They will also need the skills to build bridges with Congress: a peripatetic master, perhaps, but one who can prevent the ORI from getting kicked around. ■

➲ NATURE.COM
Discuss this article online at:
go.nature.com/da6aed

Colin Macilwain *writes on science policy from Edinburgh, UK.*
e-mail: cfmworldview@gmail.com

RESEARCH HIGHLIGHTS

Selections from the scientific literature

NEUROSCIENCE

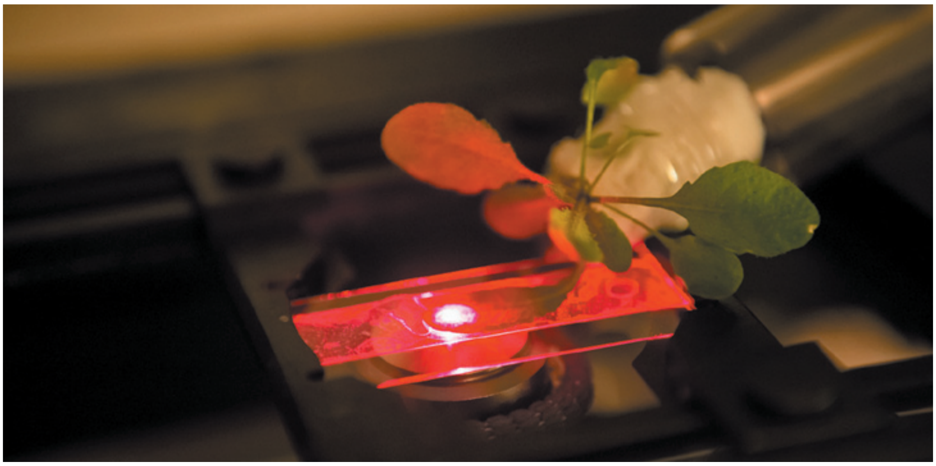
Orange light boosts brain power

The colour of the light you are exposed to could affect how well your brain functions.

Gilles Vandewalle at the University of Liège, Belgium, Howard Cooper at the French National Institute of Health and Medical Research in Bron and their colleagues exposed 16 volunteers to blue, green or orange light for 10 minutes, then kept them in darkness for 70 minutes. The researchers recorded the volunteers' brain activity while the participants performed various cognitive tests under green light. People exposed to orange light had more activity in the prefrontal cortex, which is involved in higher cognitive function, than those exposed to blue light.

The findings suggest that a light-sensitive protein in the retina called melanopsin, which is primed by orange–red light and is not involved in vision, can influence cognition.

Proc. Natl Acad. Sci. USA **http://doi.org/rwf (2014)**



BRYCE VICKMARK

BIONANOTECHNOLOGY

Nanotubes rev up photosynthesis

Carbon nanotubes that infiltrate leaves can boost photosynthetic activity and even turn plants into chemical sensors.

Michael Strano and his colleagues at the Massachusetts Institute of Technology in Cambridge used near-infrared microscopes to track single-walled carbon nanotubes in the leaves of *Arabidopsis* plants (pictured) and in extracted chloroplasts, the photosynthesizing organelle in plants. They found that the nanotubes integrate themselves into the chloroplasts' outer envelope. The semiconducting carbon nanotubes tripled photosynthetic activity in extracted chloroplasts compared with those without nanoparticles, by enhancing electron transport.

In addition, the authors showed that leaves containing fluorescent carbon nanotubes designed to stop glowing in the presence of the pollutant nitric oxide also stopped fluorescing when exposed to the pollutant. Carbon nanotubes could allow plants to detect other chemicals such as pesticides, the researchers say.

Nature Mater. **http://doi.org/rxc (2014)**

ANIMAL BEHAVIOUR

Badgers roam many miles

Badgers could be travelling much greater distances than previously supposed, a finding that might affect how the animals are culled or vaccinated to prevent the spread of bovine tuberculosis.

Andrew Byrne at University College Dublin and his colleagues marked 963 European badgers (*Meles meles*; **pictured**) between 2008 and 2012 and later recaptured them to determine their movements across 755 square kilometres of County Kilkenny in Ireland. Roughly half of the recaptures took place at the same burrows where an individual was originally captured. Of the remaining animals, 43% travelled less than 1 km. But some ventured relatively long distances: 5% of movements were over 7.3 km and one animal travelled over 22 km.

The authors suggest that officials seeking to confine disease to specific areas should extend buffer zones by at least 7.3 km.

J. Anim. Ecol. **http://doi.org/rwh (2014)**

DAVID J. SLATER/ALAMY

ECOLOGY

Killing dingoes has side effects

Efforts to control populations of a top predator in Australia have had unintended ecological effects — decreased vegetation and fewer small mammals.

In southeastern Australia, government officials have been poisoning dingoes (*Canis lupus dingo*) to reduce the predator's impact on livestock. Mike Letnic at the University of New South Wales in Sydney and his team compared seven sites in the eucalyptus forests of southeastern Australia, in which dingoes are controlled, with the same number of ecologically similar sites where the animals are left alone. Areas where dingoes are killed contained more of the animal's prey, such as kangeroos, but less understory vegetation and fewer small mammals.

As the prey populations grow in areas with dingo control, these animals consume more vegetation, reducing the cover under which small mammals can hide from their predators.

Proc. R. Soc. B 281, **20133094 (2014)**

MOLECULAR BIOLOGY

How proteins find their DNA target

Researchers have captured images of individual proteins searching for their DNA-binding sites, and have quantified parts of this process.

Zhe Liu and his team at the Howard Hughes Medical Institute in Ashburn, Virginia, looked at transcription factors — proteins that bind to specific genes to regulate their activity — in individual, living mouse embryonic stem cells. The researchers used microscopes to track single molecules of Sox2 and Oct4, key stem-cell gene regulators. They found that the two molecules use trial and error to seek out their binding sites, by colliding with DNA, diffusing away and colliding again roughly 90 times before finding their targets. The process takes about 6 minutes and the factors remain bound to the target sites for about 12 to 14 seconds.

Sox2 collides and then slides along short stretches of DNA. Sox2 also binds before Oct4, helping it to find its home.

Cell 156, **1274–1285 (2014)**

GLACIOLOGY

More of Greenland is starting to melt

Once-stable glaciers in northeast Greenland are now shrinking as a result of regional warming.

Shfaqat Khan at the Technical University of Denmark in Kongens Lyngby and his colleagues analysed satellite and airborne-elevation measurements of three glaciers that comprise the 600-kilometre-long northeast Greenland ice stream. The measurements revealed that, after more than a quarter of a century of stability, this stream began to thin between 2003 and 2006. The team then looked at ocean and atmospheric temperature data and determined that warmer air temperatures, beginning in 2003, probably led to the melting of sea ice in the region, allowing the glaciers to flow more freely into the ocean.

These ice-loss measurements exceed projections made for this region, and so models of global sea-level rise may underestimate Greenland's contribution over the coming century.

Nature Clim. Change **http://doi.org/rxb (2014)**

AGRICULTURE

Warming climate threatens crops

Climate change could result in decreasing yields of staple food crops in most parts of the world from the 2030s onwards.

Andrew Challinor at the University of Leeds, UK, and his team compared the results of more than 1,700 simulations of climate change impacts — with and without various adaptation strategies — on annual wheat, rice and maize (corn) yields. The data suggest that, without adaptation, average food-crop supplies will decline by around 5% per degree Celsius of warming.

Losses of wheat in temperate climates and rice in tropical environments can be avoided, or even reversed, by changing factors such as crop varieties, planting time, fertilizer use and irrigation. But such adaptation measures may provide little or no benefit for maize yields, particularly in tropical regions, the authors warn.

Nature Clim. Change **http://doi.org/rw9 (2014)**

HUMAN EVOLUTION

Hominin explorers were poor planners

Hominin migrations, such as those out of Africa, might have been led by indivduals with low levels of foresight.

A team led by Colin Wren at McGill University in Montreal, Canada, modelled the migratory behaviour of individuals, based on the complexity of their environments and their cognitive ability to assess and seek out better environments.

The model predicted that those with poorer foresight were more likely to stumble on a better habitat, and therefore disperse into new areas. Furthermore, homogeneous environments fostered exploration because the more constant level of resources allowed explorers to travel further than those in a more heterogeneous setting.

J. Hum. Evol. **http://doi.org/rwd (2014)**

COMMUNITY CHOICE

The most viewed papers in science

NANOTECHNOLOGY

3D printing goes nanoscale

HIGHLY READ on pubs.acs.org in February

Researchers have used a three-dimensional (3D) printing technique to stack nanometre-scale polymer threads on top of one another, forming a tiny wall.

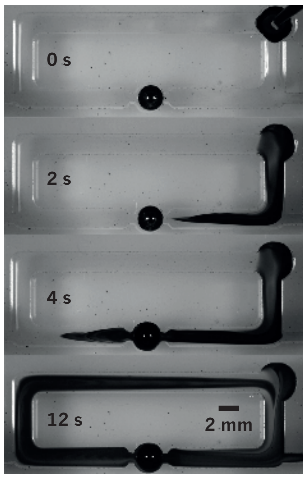
A promising 3D nanoscale printing method involves spinning nanofibres out of a liquid under a strong electrical field. But only certain shapes have been printed this way, because the fibres tend to electrically repel each other. To solve this problem, Minhee Lee and Ho-Young Kim of Seoul National University fired polymer nanofibres onto a metal template sitting on top of a grounding plate that quickly drains the deposited fibres of charge. The fibres then spontaneously attract each other, and stack to form a free-standing wall.

The technique cannot yet be used to print objects with sharp corners but could still be useful for building, for example, tiny transistors, the authors say.

Langmuir 30, **1210–1214 (2014)**

PHYSICS

Micro-pump with no moving parts



PHRED PETERSEN/SHIYANG TANG/RMIT

A pump can propel liquids along millimetre-sized channels at high speeds with no mechanical parts.

The pump, designed by Khashayar Khoshmanesh at the Royal Melbourne Institute of Technology University in Australia and his colleagues, is a droplet of metal — an alloy of gallium, indium and tin — that is held loosely in a spherical chamber in the middle of a Plexiglas channel. A slow chemical reaction in the metal droplet results in gallate anions diffusing to the droplet's surface. When a low-power electric field is applied, the rearrangement of charges at the droplet's surface causes it to shift in its constrained space in such a way that the surrounding solution flows past it (**pictured**).

The pump might be useful in microscopic machines, the researchers say.

Proc. Natl Acad. Sci. USA 111, **3304–3309 (2014)**

NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

POLICY

Europe HFC ban

The European Parliament has approved a ban on potent greenhouse gases used in some cooling systems, aerosol sprays and synthetic foams. Once in the atmosphere, the warming effect of hydrofluorocarbon (HFC) gases is thousands of times that of carbon dioxide, and emissions of HFCs across Europe have increased by 60% since 1990. The legislation aims to cut use of the gases by 79% over the next 15 years. European Union member states will vote on the new rules in April.

GM crops detected

Small amounts of genetically modified (GM) crops are increasingly being detected in traded food and feed, says the Rome-based United Nations Food and Agricultural Organization (FAO). In the first survey of its kind, published on 13 March, the FAO reports that, across 75 of its member countries, regulators identified 198 incidents in which low levels of GM crops were detected in supposedly non-GM crops. Most cases occurred in the latter years of the survey between 2009 and 2012, and involved linseed, rice, maize (corn) and papaya. Once detected, most shipments were destroyed or returned to the country of origin. The FAO attributes the rise to increased global production of GM crops and better detection technology.

AZWAR/EPA/CORBIS

Illegal fires ravage Sumatra

The number of fires raging across the Indonesian island of Sumatra (pictured) last week is the highest in recent years. Using satellite data, the World Resources Institute (WRI) in Washington DC detected 3,101 fires on the island. This exceeds the 2,643 fires detected in the same region in June last year, when the nation's government declared a state of emergency in response to the worst air pollution ever recorded in southeast Asia. Roughly half of the current fires are on land managed by oil-palm and logging companies, even though using fire to clear land is illegal in Indonesia, a WRI analysis says.

Dangerous research

Animal-rights extremists are increasingly making their attacks personal. A 12 March report by the Federation of American Societies for Experimental Biology says that since 2000, nearly 50% of attacks by extremists have targeted individuals rather than institutions, compared to 9% over 1990–99. Extremists are harassing researchers at home and destroying their property, the report says. It recommends that researchers protect themselves by becoming familiar with local laws and keeping their personal contact information private.

Undefined illness

The US Institute of Medicine cannot define Gulf War illness, which plagues veterans of the 1990–91 war with symptoms of fatigue, pain, memory loss and gastrointestinal disorders. The diversity of symptoms and the lack of a diagnostic test prevented a single definition, said a report published on 12 March by the institute, based in Washington DC. The Department of Veterans Affairs had asked for a more specific definition to guide research and care for veterans. The department argues that the condition should not be treated primarily as a mental illness.

Swiss grants crisis

The Swiss government has stepped in to offer substitute grants in the wake of a move that bars researchers in the country from applying to the European Research Council (ERC). Switzerland lost its status as an associate partner in the European Union's €80-billion (US$110-billion) Horizon 2020 funding programme after it imposed curbs on immigration (see *Nature* **506,** 277; 2014). The change hit hundreds of would-be ERC applicants based in the country. The Swiss National Science Foundation in Bern is now offering a number of replacement grants of comparable size to those offered by the ERC. See go.nature.com/bfldcw for more.

Climate consensus

The American Association for the Advancement of Science issued a report on 18 March called *What We Know*, summarizing the science behind climate change (see go.nature.com/ewy8cd). It follows an overview of climate science by the UK Royal Society and the US National Academy of Sciences that was

launched on 27 February (see go.nature.com/puvn4v). The documents come ahead of the Intergovernmental Panel on Climate Change's report focusing on the impacts of climate change, which will be released at the end of March.

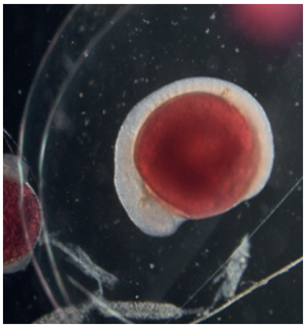
RESEARCH

Inflation evidence

A telescope at the South Pole has revealed strong evidence that the Universe went through a period of rapid inflation just after the Big Bang. To great excitement, a collaboration led by researchers at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, announced that the telescope had spotted the influence of gravitational waves (ripples in space that inflation would have caused) on the cosmic microwave background, the radiation released after the Big Bang. See page 281 and go.nature.com/lruz8e for more.

Carp invasion

Asian carp that have run amok in the rivers of the midwestern United States are spawning farther north in the Mississippi River than previously recorded, the US Geological Survey (USGS) said on 11 March. USGS scientists found embryos (**pictured**) of bigheaded carp about to hatch in the river running through Lynxville, Wisconsin. The fish were originally imported from southeast Asia to the southern United States for aquaculture and to help clean ponds in wastewater-treatment facilities. But they escaped and bred, and are now causing environmental and economic damage.



AMY GEORGE/USGS

Research apology

The Japanese research institute that is home to several authors of two controversial stem-cell papers has apologized for errors in the research. On 14 March, the RIKEN institute in Tokyo announced the preliminary findings of an investigation into work led by Haruko Obokata that describes a method for reprogramming differentiated mouse cells into a pluripotent embryonic-like state by stressing them. The papers, published in *Nature*, were criticized for using several duplicated images and because their results have not been reproduced. See page 283 for more.

PEOPLE

Córdova confirmed

The US Senate confirmed astrophysicist France Córdova as head of the National Science Foundation (NSF) on 12 March. The foundation funds basic non-medical research with a budget of US$7.2 billion; Córdova's confirmation comes just over a year after former director Subra Suresh resigned mid-term. Córdova was most recently chairwoman of the Smithsonian Institution's Board of Regents and a member of the National Science Board, the panel that oversees the NSF. See page 285 for more.

Integrity red tape

The departing director of the US Office of Research Integrity (ORI), David Wright, has accused the Department of Health and Human Services — in which the ORI sits — of "remarkably dysfunctional" bureaucracy in a resignation letter. Wright, who is leaving the directorship on 27 March after two years in charge, described what he said was the low priority afforded his office by higher management; for example, the ORI's head-of-education post has been unfilled for 16 months. The letter was published online by *Science* on 12 March. Wright said that bureaucracy "sucks away time and resources that we might better use to meet our mission". See page 275 for more.

COMING UP

25–28 MARCH
Marking the 100th anniversary of the birth of Norman Borlaug, who saved millions from starvation by developing higher-yielding crops, the International Maize and Wheat Improvement Center in Mexico is holding a conference to discuss the state of research on wheat.
go.nature.com/hrne9g

26–28 MARCH
Physicists debate a suitable landing site for the ExoMars rover at a meeting at the European Space Agency's European Space Astronomy Centre near Madrid. The mission plans to land a rover on the red planet in 2018.
go.nature.com/i5n5r2

BUSINESS

BP drilling back on

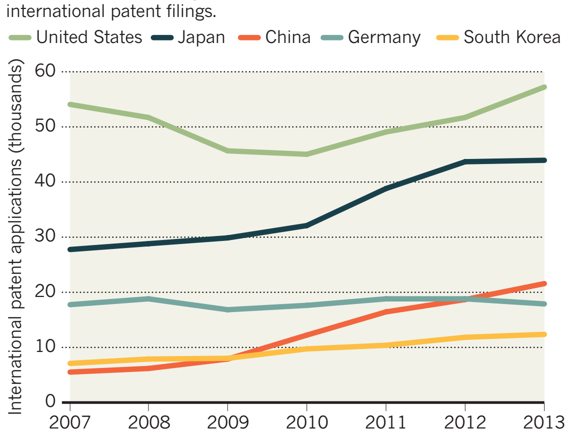
Energy giant BP will once more be allowed to bid for US leases to drill for oil and gas offshore, after the company reached an agreement with the US Environmental Protection Agency (EPA) on 13 March. It had been barred from new federal contracts in November 2012, two years after its Deepwater Horizon drilling rig in the Gulf of Mexico was involved in a massive oil spill. The agreement comes with conditions — for example, BP must improve its practices and an EPA-approved auditor will review them annually.

TREND WATCH

International patent applications grew 5.1% to more than 200,000 last year, according to the World Intellectual Property Organization in Geneva, Switzerland. The United States exceeded its previous record, set before the financial crisis, and China saw the largest growth with a 16% surge. The University of California (46% rise to 398 applications) and the Massachusetts Institute of Technology in Cambridge (49% rise to 217) were the top-filing educational institutions.

CHINA'S PATENTS SURGE

China overtook Germany in 2013 to reach third place in international patent filings.



SOURCE: WIPO

NATURE.COM
For daily news updates see:
www.nature.com/news

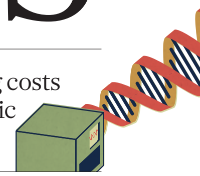
NEWS IN FOCUS

BIOLOGY Acid-bath stem-cell study found to contain 'serious errors' **p.283**

POLICY Astrophysicist approved as the head of US National Science Foundation **p.285**

ECOLOGY Dam release set to flood Colorado River with new life **p.286**

GENOMICS Sequencing costs get close to the magic number **p.294**



STEFFEN RICHTER/HARVARD UNIVERSITY

The BICEP2 instrument (foreground) at the South Pole has detected signs of ripples from the Universe's first moments.

COSMOLOGY

Telescope captures view of gravitational waves

Images of the infant Universe reveal evidence for rapid inflation after the Big Bang.

BY RON COWEN IN CAMBRIDGE, MASSACHUSETTS

Astronomers have peered back to nearly the dawn of time and found what seems to be the long-sought 'smoking gun' for the theory that the Universe underwent a spurt of wrenching, exponential growth called inflation during the first tiny fraction of a second of its existence.

Using a radio telescope at the South Pole, the US-led team has detected the first evidence of primordial gravitational waves, ripples in space that inflation generated 13.8 billion years ago when the Universe first started to expand.

The telescope captured a snapshot of the waves as they continued to ripple through the Universe some 380,000 years later, when stars had not yet formed and matter was still scattered across space as a broth of plasma. The image was seen in the cosmic microwave background (CMB), the glow that radiated from that white-hot plasma and that over billions of years of cosmic expansion has cooled to microwave energies.

The fact that inflation, a quantum

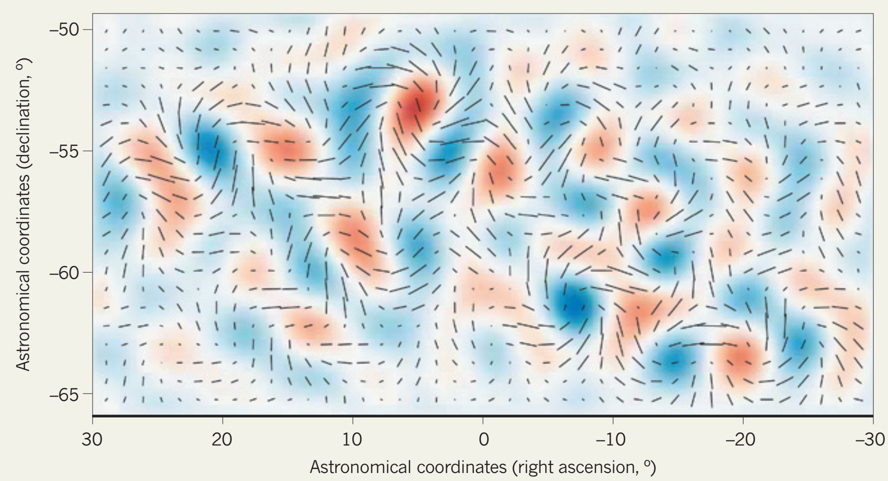
NATURE.COM
For more on cosmic ripples from the Big Bang:
go.nature.com/cjjh5y

phenomenon, produced gravitational waves demonstrates that gravity has a quantum nature just like the other known fundamental forces of nature, experts say. Moreover, it provides a window into interactions much more energetic than are accessible in any laboratory experiment. In addition, the way that the team confirmed inflation is itself of major significance: it is the most direct evidence yet that gravitational waves — a key but elusive prediction of Albert Einstein's general theory of relativity — exist.

"This is a totally new, independent piece of cosmological evidence that the inflationary picture fits together," says theoretical physicist Alan Guth of the Massachusetts Institute of Technology (MIT) in Cambridge, who proposed the idea of inflation in 1980. He adds that the study is "definitely" worthy of a Nobel prize.

COSMIC CURL

The BICEP2 instrument observed a faint but distinctive twisting pattern, or spin, known as a curl or B-mode, in the polarization of the cosmic microwave background. This is the first evidence for gravitational waves generated by rapid inflation of the Universe some 13.8 billion years ago.



SOURCE: BICEP2 COLLABORATION

INSTANT INFLATION

Guth's idea was that the cosmos expanded at an exponential rate for a few tens of trillionths of trillionths of trillionths of seconds after the Big Bang, ballooning from subatomic to football size. Inflation solves several long-standing cosmic conundrums, such as why the observable Universe appears uniform from one end to the other. Although the theory has proved to be consistent with all cosmological data collected so far, conclusive evidence for it has been lacking.

Cosmologists knew, however, that inflation would have a distinctive signature: the brief but violent period of expansion would have generated gravitational waves, which compress space in one direction while stretching it along another (see 'Ripple effect'). Although the primordial waves would still be propagating across the Universe, they would now be too feeble to detect directly. But they would have left a distinctive mark in the CMB: they would have polarized the radiation in a curly, vortex-like pattern known as the B mode (see 'Cosmic curl').

Last year, another telescope in Antarctica — the South Pole Telescope (SPT) — became the first observatory to detect a B-mode polarization in the CMB (see *Nature* http://doi.org/rwt; 2013). That signal, however, was over angular scales of less than one degree (about twice the apparent size of the Moon in the sky), and was attributed to how galaxies in the foreground curve the space through which the CMB travels (D. Hanson *et al. Phys. Rev. Lett.* **111,** 141301; 2013). But the signal from primordial gravitational waves is expected to peak at angular scales between one and five degrees.

And that is exactly what John Kovac of the Harvard-Smithsonian Center for Astrophysics (CfA) in Cambridge, Massachusetts, and his colleagues now say they have detected, using an instrument dubbed BICEP2 that is located just metres away from its competitor, the SPT.

Detecting the tiny B mode required measuring the CMB with a precision of one ten-millionth of a kelvin and distinguishing the primordial effect from other possible sources, such as galactic dust.

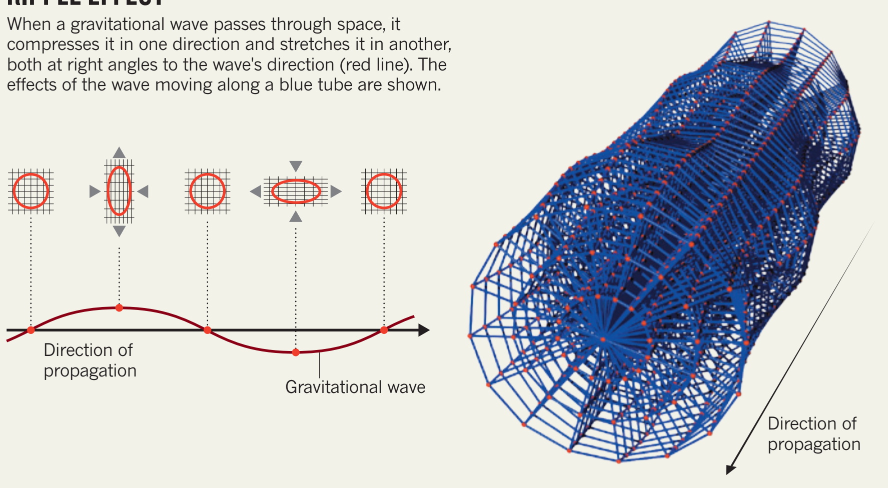
"The key question," says Daniel Eisenstein, an astrophysicist at the CfA, "is whether there could be a foreground that masquerades like this signal". But the team has all but ruled out that possibility, he says. First, the researchers were careful to point BICEP2 — an array of 512 superconducting microwave detectors — at the Southern Hole, a patch of sky that is known to contain only tiny amounts of such emissions. They also compared their data with those taken by an earlier experiment, BICEP1, and showed that a dust-generated signal would have had a different colour and spectrum.

Furthermore, data taken with a newer, more sensitive polarization experiment, the Keck array, which the team finished installing at the South Pole in 2012 and will continue operating for two more years, showed the same characteristics. "To see this same signal emerge from two other, different telescopes was for us very convincing," says Kovac.

"The details have to be worked out, but from what I know it's highly likely this is what we've all been waiting for," says astronomer John Carlstrom of the University of Chicago, Illinois, who is the lead researcher on the SPT. "This is the discovery of inflationary gravitational waves."

RIPPLE EFFECT

When a gravitational wave passes through space, it compresses it in one direction and stretches it in another, both at right angles to the wave's direction (red line). The effects of the wave moving along a blue tube are shown.



SOURCE: MARKUS PÖSSEL/EINSTEIN-ONLINE.INFO

SOLID SIGNATURE

Cosmologist Marc Kamionkowski adds: "To me, this looks really, really solid." He was one of the first cosmologists to calculate what the signature of primordial

gravitational waves should look like in the CMB. The findings are "on a par with dark energy, or the discovery of the CMB — something that happens once every several decades", says Kamionkowski, who is at Johns Hopkins University in Baltimore, Maryland.

The strength of the signal measured by BICEP2, although entirely consistent with inflation, initially surprised the researchers because it is nearly twice as large as estimated from previous experiments. According to theory, the intensity of a B-mode signal reveals how fast the Universe expanded during inflation, and therefore suggests the energy scale of the cosmos during that epoch. The data pinpoint the time when inflation occurred — about 10^{-37} seconds into the Universe's life — and its temperature at the time, corresponding to energies of about 10^{16} gigaelectronvolts, says cosmologist Michael Turner of the University of Chicago. That is the same energy at which three of the four fundamental forces of nature — the weak, strong and electromagnetic force — are expected to become indistinguishable from one another in a model known as the grand unified theory.

Because inflation took place in the realm of quantum physics, seeing gravitational waves arise from that epoch provides "the first-ever experimental evidence for quantum gravity", says MIT cosmologist Max Tegmark — in other words, it shows that gravity is at heart a quantum phenomenon, just like the other three fundamental forces. Physicists, however, have yet to fully understand how to reconcile general relativity with quantum physics from a theory standpoint.

The researchers reported the findings on 17 March at a press briefing at the CfA, held just after they described their results to scientists in a technical talk. The team also released several papers describing the results. In so doing, it seems to have beaten the SPT and also several other groups racing to find the fingerprint of inflation using an assortment of balloon-borne and ground-based experiments and one satellite, the European Space Agency's Planck spacecraft.

More-extensive maps of the B-mode polarization, and especially a full-sky survey, which the Planck telescope may be able to obtain later this year, should provide more clues about how inflation unfolded and what drove it. In addition to looking farther back in time than ever before, the discovery "is opening a window a trillion times higher in energy than we can access with the Large Hadron Collider", the world's premiere atom smasher, notes cosmologist Avi Loeb of the CfA, who is not part of the BICEP2 team. ■

CELL BIOLOGY

Stem-cell method faces fresh questions

Papers describing acid-bath technique under more scrutiny after institute's investigation finds errors in methodology.

BY DAVID CYRANOSKI

The veracity of two papers that detailed a method to reprogram mature cells into an embryonic state by exposing them to stress has come under more pressure. Days after the first author's institute reported "serious errors" in the papers' methodology, questions were raised over the same researcher's doctoral dissertation and the cells used in the study.

On 14 March, RIKEN, Japan's largest research organization, which runs the Center for Developmental Biology (CDB) in Kobe where first author Haruko Obokata and several of her co-authors work, announced the interim findings of its investigation into allegations of irregularities in the method. One RIKEN investigator advised them to retract the papers. Then, in further developments, Obokata called into question the quality of her own doctoral thesis, which is already under investigation by the university that granted it. And a co-author on the reprogramming papers says that he is sending some of the cells produced in the experiments for independent verification.

The two papers made headlines around the world when they were published in *Nature* on 30 January (H. Obokata *et al. Nature* **505,** 641–647 and 676–680; 2014). The technique, which the team called stimulus-triggered acquisition of pluripotency (STAP), is important because cells reprogrammed into an embryonic state are ideal for studying the development of disease or the effectiveness of drugs.

But within weeks, the papers were attacked by scientists over their use of several duplicated images and by those who could not reproduce the work, prompting RIKEN to investigate.

At the press conference, a five-person panel, including RIKEN director and Nobel-prize-winner Ryoji Noyori, noted six problems. Two were dismissed as unintentional mistakes. Four others — including an image of an electrophoresis gel that seemed to have had a lane added in later, and plagiarism in part of the methods section — were deemed more serious and are still under investigation. The panel offered no clear answers about whether the STAP phenomenon is real, but noted that RIKEN co-author Hitoshi Niwa was attempting to replicate the method. It added that there were no signs of fraud.

Panel member Masatoshi Takeichi, director of the CDB, also told reporters that the three co-authors from the institute — Obokata, Niwa and Yoshiki Sasai — had agreed to retract the paper at his request. But a statement in Japanese by the trio, delivered at the press conference, said only that they were "considering a retraction and contacting outside authors to discuss that possibility". Charles Vacanti, of Harvard Medical School in Boston, Massachusetts, who is the senior corresponding author on the first of the two papers (H. Obokata *et al. Nature* **505,** 641–647; 2014), has made it clear that he has no intention of retracting unless there is compelling evidence that the data are incorrect.

To add confusion, the first 20 pages of Obokata's thesis, completed in 2011 at Waseda University in Tokyo, were found by Nature News to be taken from a US National Institutes of Health primer on stem cells, and one image in the results section has been reproduced from a commercial website without a citation. Moreover, Vacanti, who was listed on the thesis as a member of the examination committee that approved it, told Nature News: "I was not presented with or asked to read a copy of her dissertation."

"What did I inject into those blastocysts? This is what I want to know."

Last week, Obokata wrote to an unnamed professor at Waseda University indicating that she wanted to retract the thesis. She has not, however, formally requested a retraction.

A desire to resolve the STAP controversy has led to an investigation into the identity of the cells in the papers. Teruhiko Wakayama of the University of Yamanashi, a senior author on the second paper (H. Obokata *et al. Nature* **505,** 676–680; 2014), helped to test the pluripotency of Obokata's STAP stem cells by injecting them into mouse embryos. By turning into different cell types within these mice, the cells proved they had the developmental capacity that STAP promises. But he has now sent the cells that Obokata gave him to an independent institute for genetic analysis to see if they are really STAP cells. "What did I inject into those blastocysts?" Wakayama asks. "This is what I want to know more than anything else." He hopes to have his answer in the next few months. ■

Nature News is editorially independent of the research publications section of *Nature*.

POLICY

Incoming NSF director faces challenges in Congress

Former Purdue University president France Córdova inherits an agency at a crossroads.

BY JESSICA MORRISON

By all accounts, astrophysicist France Córdova thrives in the face of a challenge. As NASA's chief scientist in the mid-1990s, she helped then-administrator Daniel Goldin to push his 'faster, better, cheaper' vision for space research. Later, as chancellor of the University of California, Riverside, she won approval for a long-sought medical school. And during her 2007–12 tenure as president of Purdue University, she helped to double research funding despite a global recession.

That persistence should serve her well as director of the US National Science Foundation (NSF), the position she assumed after her appointment was confirmed by the Senate on 12 March. The US$7.2-billion agency, which funds basic non-medical research, has seen its budget stagnate in recent years and is now facing attacks on its peer-review system and social-science division from conservative members of Congress.

"The toughest challenge that she's facing right now is one that is faced by all of the research institutions that are funded by taxpayer dollars — relevance," says Internet pioneer Vinton Cerf, vice-president of Google and a member of the National Science Board, which oversees the NSF.

POLITICAL EXPERIENCE

Córdova studied English at Stanford University in California, and received a physics doctorate from the California Institute of Technology in Pasadena in 1979. She then spent a decade at Los Alamos National Laboratory in New Mexico, studying X-ray and γ-ray sources, before making her first foray into administration at Pennsylvania State University in University Park, where she served as head of the astronomy department from 1989 to 1993. Her subsequent moves to NASA, the University of California, Riverside, and Purdue in West Lafayette, Indiana, saw her take on positions of increasing responsibility.

Córdova is no stranger to the NSF or its struggles. Until her confirmation, she was a member of the National Science Board, and chaired the governing board of the Smithsonian Institution in Washington DC, the world's largest museum complex. Her time on the science board is likely to shorten her learning curve as she assumes the NSF's top job, says environmental microbiologist Rita Colwell, who led the agency from 1998 to 2004.

NSB

France Córdova became head of the National Science Foundation on 12 March.

Goldin, now chief executive of the Intellisis Corporation in San Diego, California, says that Córdova has already proved her ability to operate in the thorny political environment of Washington DC. At NASA — her first government post — she was pivotal to implementing Goldin's controversial plan to expand the agency's scientific reach and stretch its budget by backing smaller, less expensive science and exploration missions. "She developed really good, solid relationships with staff in the House and Senate, and the White House," he says. "She is able to communicate with people on the issues and not emotions."

That experience, and her subsequent work as a university administrator, is not far from Córdova's mind as she assumes her new role. "I don't feel like I've ever left government," she told reporters the day after her Senate confirmation. "I've been involved in one way or another — taking money or giving it."

With austerity a priority in Washington DC, Córdova is already thinking about how to make the NSF's research dollars go further. "Budgets are limited, and that's always a challenge," she tells *Nature*. "But I think that government organizations like NSF could work toward more partnerships with industry, foundations and private donors, to leverage the amount of money that each has."

CONGRESSIONAL CAMPAIGN

Córdova has a long-term goal to further the foundation's emphasis on interdisciplinary and international research programmes — a focus of her immediate predecessor as NSF director, engineer Subra Suresh, who resigned a little over a year ago. But she also faces a more immediate challenge: winning over some of her agency's staunchest critics, members of Congress who are questioning the quality of the research that the NSF supports.

On 10 March, Republican members of the House of Representatives' science committee introduced a bill that would require the NSF to certify each grant it awards as serving the "national interest" in one of six categories (see *Nature* http://doi.org/rxg; 2013). The bill, put forward by congressmen Lamar Smith (Republican, Texas) and Larry Bucshon (Republican, Indiana), also seeks to slash funding for the NSF's social-science directorate by 22%, to $200 million, while increasing spending on many of the agency's other major research divisions. That would commit more of the NSF's limited funds to producing discoveries that drive economic growth, says Smith.

> ***"We have to better explain why we do what we do."***

The debate over the bill in the House and the science community echoes a simple sentiment that Córdova heard over and over again in recent weeks as she spent time getting to know key members of Congress and their aides while waiting for the Senate to confirm her new post. "We have to better explain why we do what we do," she says of the NSF.

Córdova's operating style may be just what is needed to get the message across. Former colleagues say that she draws on her diplomacy and innate curiosity to build a broad base of support for her goals. "She surrounds herself with clear thinkers, and she solicits their opinions," says Philip Low, a chemist who directs Purdue's Center for Drug Discovery. "No one is the sole repository of wisdom, and she recognizes that." ■

PETE MCBRIDE/NATL GEOGR. SOC./CORBIS

Water diversions in two countries mean that the mighty Colorado River peters out before it reaches Mexico's Gulf of California.

ECOLOGY

Water returns to arid Colorado River delta

US–Mexico agreement paves the way for a rare environmental test.

BY ALEXANDRA WITZE

One of North America's most iconic rivers is about to undergo an unprecedented experiment in ecological engineering.

On 23 March, operators at the Morelos Dam along the US–Mexico border near Yuma, Arizona, will open the gates and begin releasing water downstream. The goal is to dampen broad swathes of the arid Colorado River delta for the first time in decades, allowing new cottonwood and willow trees to germinate and restore small patches of riparian habitat.

The move, which follows bitter international battles over water rights, will mark the first time that the United States and Mexico have put water back into the parched riverbed for environmental purposes. It is both a practical and a symbolic victory for conservationists who have fought to restore what was once 800,000 hectares of lush wetlands, as well as a rare opportunity for ecologists worldwide to watch what happens.

The mighty Colorado rises on the western slope of the Rocky Mountains and drains seven US and two Mexican states along its 2,300-kilometre course (see 'River run'). Before the 1930s, when dams began to throttle the river, its water ran unfettered into the Gulf of California. But most was soon diverted to slake the thirst and agricultural fields of millions of people in the American Southwest. A 1944 international treaty granted Mexico just 10% of the river's original flow; vast delta wetlands shrivelled to patches of vegetation clinging to sandy plains.

In 2012, officials drew up an addendum to the original water treaty. Known as Minute 319, the agreement lays out how the United States and Mexico will share water surpluses and shortages until the end of 2017. But it also mandates the experimental release of what it calls "water for the environment".

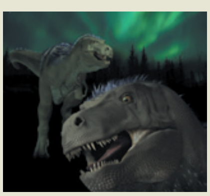
That was the call to action for a group of researchers from universities, government agencies and non-governmental organizations who had been working for years on delta conservation (E. P. Glenn, K. W. Flessa and J. Pitt *Ecol. Eng.* **59,** 1–6; 2013). They calculated how much water should be released, and over what period of time. As a result of their recommendations, over nearly eight weeks, dam operators will allow some 130 million cubic metres of water to travel downstream.

That is less than 1% of the river's average annual flow, but it is still significant, says Francisco Zamora Arroyo, director of the Colorado River Delta Legacy Program at the

MORE ONLINE

KAREN CARR

DINO OF THE WEEK



- *T. rex* adapted to Arctic climate by shrinking in size go.nature.com/3xylun

MORE NEWS

- How 'glory' seen on Venus differed from a rainbow go.nature.com/tfpyrh
- Beetle now impervious to Bt toxin in transgenic maize go.nature.com/obmzzz
- Reindeer spooked by power lines' ultraviolet flashes go.nature.com/mzoge7

NATURE PODCAST

Mother's vitamin A deficiency affects baby's immunity through life; and gravitational waves from inflation nature.com/nature/podcast

Sonoran Institute in Tucson, Arizona. "It's going to allow us to learn from this event and better determine the water needs for the delta in the future," he says.

Most of the flow will come in a three-day pulse, from 27 to 29 March, which is meant to mimic a small-scale spring flood (K. W. Flessa *et al. Eos* **94,** 485–486; 2013). "It's structured so that we hopefully get a big pulse crossing over into places that have a lot of restoration potential," says team co-leader Karl Flessa, a geoscientist at the University of Arizona in Tucson. "And we hope that the groundwater doesn't recede too quickly after that."

In the days and weeks after the water pulse, monitoring teams will descend on the delta, because no one is entirely sure how long the water will last or what its exact course will be. Hydrologists will gather data on changes in surface water and groundwater; ecologists will count saplings and determine whether that new habitat is attracting birds.

The experiment could benefit the almost 400 bird species that live in the delta, says Osvel Hinojosa Huerta, an ecologist with Pronatura Noroeste, a conservation group with offices in San Luis Río Colorado, Mexico. Every year, about 400,000 migratory waterbirds also pass through the region, an important stop on their flyway along the Americas.

In recent decades, invasive saltcedar trees have crowded out much of the native habitat. The water pulse is designed to create broad expanses of moist sand in which native cottonwoods and willows can germinate, says Patrick Shafroth, a plant ecologist with the US Geological Survey in Fort Collins, Colorado.

The challenge will be to keep those newborn trees alive. Shafroth helped to lead related experiments on a Colorado River tributary, the Bill Williams River in Arizona, in 2005 and 2006. He says that timing is critical for getting seeds distributed to sites where they can take root and grow quickly enough to survive. Thanks in part to such lessons, the delta pulse flow is designed to try to keep the water table from dropping by more than 2.5 centimetres per day, the rate at which cottonwood roots can keep up.

In addition to the eight-week pulse flow, dam operators will also release a 'base flow' of 65 million cubic metres, on a schedule that is yet to be determined. The idea is to try to nourish and maintain any new growth that appears this spring.

Team leaders acknowledge that plenty of things could go wrong. Water might vanish into unexpected places, seeping deep underground where it does little good. Trees might not germinate if their seeds do not spread properly. And many saplings could perish if the summer is particularly brutal. But a successful experiment could build support for future releases to the delta, although only one is provided for by the Minute 319 agreement.

"We hope that the groundwater doesn't recede too quickly."

Whatever happens during the experiment will almost certainly come as a surprise, says John Schmidt, an ecologist at Utah State University in Logan, who also heads the US Geological Survey's Grand Canyon Monitoring and Research Center in Flagstaff, Arizona. In 1996, Schmidt helped to lead some larger engineered floods much higher up the Colorado. The goal was to restore sandbars and beaches in areas such as Grand Canyon National Park, but early attempts failed because the newborn beaches quickly eroded (see *Nature* **420,** 356–358; 2002).

"We don't really know what's going to happen," says Flessa. "If we did, we wouldn't be doing the experiment." ■

Thomas Insel wants studies to identify the biological mechanisms that underlie psychiatric symptoms.

LEXEY SWALL/THE NEW YORK TIMES/REDUX/EYEVINE

POLICY

NIH rethinks psychiatry trials

Mental-health division will no longer fund research aiming to relieve symptoms without probing underlying causes.

BY SARA REARDON

Thomas Insel, the director of the US National Institute of Mental Health (NIMH), has had enough of shooting in the dark. He thinks that if a clinical trial of a psychiatric therapy fails, scientists should at least learn something about the brain along the way.

Now Insel is translating that belief into action: the NIMH, based in Bethesda, Maryland, has decided to stop funding clinical trials that aim merely to ease patients' symptoms. "Future trials will follow an experimental medicine approach in which interventions serve not only as potential treatments, but as probes to generate information about the mechanisms underlying a disorder", he wrote in a 27 February blog post announcing the move. This funding switch, which will affect grants due to be made in a few months' time, intensifies the NIMH's apparent shift in emphasis from abstract psychiatry to the neurobiological roots of disease.

"It's a totally new departure for us," says Bruce Cuthbert, a clinical psychologist and director of the institute's adult translational-research division. Insel notes that the NIMH spent about US$100 million on clinical trials in 2013, and says that more than half of recipient projects received funding without any requirement to examine the biological processes involved in a disease. In many cases, "if you get a negative result you have no idea why, and you have to try something else at random", Cuthbert says. "It's an incredible waste of money."

The new rules, which will apply to the grant cycle that begins in June, also seek to increase transparency by requiring faster online registration of trials and stricter guidelines for reporting results. Insel acknowledges that researchers may have to rework their studies to satisfy the new guidelines. "I think this will be really unpopular," he says.

History shows that Insel is not daunted by controversy. In April 2013, he raised eyebrows when he announced that the NIMH would stop using the *Diagnostic and Statistical Manual of Mental Disorders 5* (*DSM-5*), the latest version of the American Psychiatric Association's diagnostic guide, to classify mental disorders (see *Nature* http://doi.org/rvd; 2013). The book's definitions tend to lump patients together by symptoms, which often do not precisely map to what is wrong with their brains, he says. With this haphazard approach to trials, even if symptoms are alleviated, researchers still may not understand what caused them. "We've studied drugs, not disorders — if you throw something at the wall, and *P* is less than 0.05, you win," he says. Such thinking, he adds, wastes time and money.

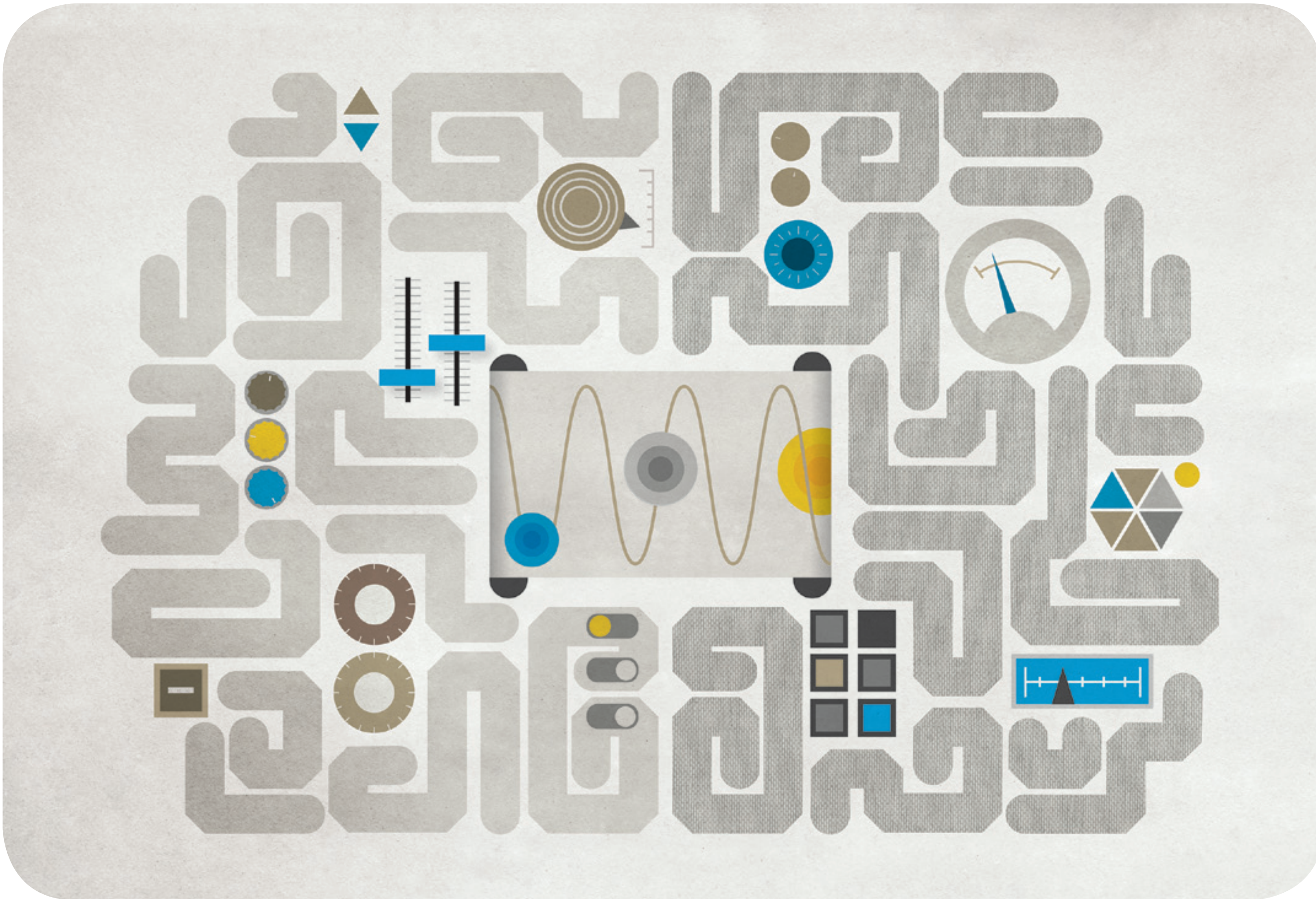
Michael First, a psychiatrist at Columbia University in New York, says that targeting the wrong patient groups can also make therapies seem less effective than they really are. A treatment that helps a tiny fraction of people with a *DSM* diagnosis of depression, for example, might be highly effective for all people with a more specific trait, such as the inability to enjoy life. Beginning a trial with an idea about a treatment's mechanism may allow researchers to quickly determine whether to expand or end the testing, First says. That could help to tempt drug companies to develop new classes of drugs for mental illnesses — an endeavour that many have abandoned after decades of failed research.

But Harold Pincus, a psychiatrist also at Columbia University, cautions that the data-driven approach is not a sure thing. Focusing trials on more specific biological targets will require researchers to think very carefully about how they recruit and classify their research subjects. Neuroscientist Michael Owen of Cardiff University, UK, concurs. "It's a pretty scary prospect for everyone if you throw away the diagnostic rule book," he says.

Others are strongly opposed to the NIMH's new approach. Psychiatrist Allen Frances, an emeritus professor at Duke University in Durham, North Carolina, worries that an emphasis on brain science comes at patients' expense. "Don't bet everything you have on something so difficult," he says. "We have a shameful problem of neglect of existing patients and an empty promissory note to the future that science will solve everything." Understanding the workings of the brain is, of course, a fine goal, he says, but "it will be decades and decades before anyone will be helped by this research".

Insel maintains that the NIMH wants the trials it funds to continue to ameliorate patients' symptoms, but to probe how their brains work at the same time. His emphasis on biological mechanisms is not unique at the US National Institutes of Health, the NIMH's parent agency: the National Institute of Neurological Disorders and Stroke in Bethesda is increasingly aiming therapies at specific targets in the brain rather than abstract constellations of symptoms. "We've all been burnt badly by treating the brain as a black box," says that agency's deputy director, neurologist Walter Koroshetz.

The difference, Koroshetz adds, is that a great deal is already known about how neurological disorders such as Parkinson's disease work, and that allows scientists to focus on other challenges, such as adapting drugs that work in animals to work in humans. Psychiatry has a long way to go before it will have such accurate targets. "I feel bad for them," Koroshetz says. ■ SEE EDITORIAL P.273



Tuning THE BRAIN

Deep brain stimulation has shown promise in treating conditions such as Parkinson's disease. Now scientists are using the technology to eavesdrop on problem neural circuits.

BY HELEN SHEN

For Frank Donobedian, sitting still is a challenge. But on this day in early January, he has been asked to do just that for three minutes. Perched on a chair in a laboratory at Stanford University in California, he presses his hands to his sides, plants his feet on the floor and tries with limited success to lock down the trembling in his limbs — a symptom of his Parkinson's disease. Only after the full 180 seconds does he relax.

Other requests follow: stand still, lie still on the floor, walk across the room. Each poses a similar struggle, and all are watched closely by Helen Bronte-Stewart, the neuroscientist who runs the lab.

"You're making history," she reassures her patient.

"Everybody keeps saying that," replies the 73-year-old Donobedian, a retired schoolteacher, with a laugh. "But I'm not doing anything."

"Well, your brain is," says Bronte-Stewart.

Like thousands of people with Parkinson's before him, Donobedian is being treated with deep brain stimulation (DBS), in which an implant quiets his tremors by sending pulses of electricity into motor areas of his brain. Last October, a team of surgeons at Stanford threaded the device's two thin wires, each with four electrode contacts, through his cortex into a deep-seated brain region known as the subthalamic nucleus (STN).

But Donobedian's particular device is something new. Released to researchers in August 2013 by Medtronic, a health-technology firm in Minneapolis, Minnesota, it is among the first of an advanced generation of neurostimulators that not only send electricity into the brain, but can also read out neural signals generated by it. On this day, Bronte-Stewart and her team have temporarily turned off the stimulating current and are using some of the device's eight electrical contacts to record abnormal neural patterns that might correlate with the tremors, slowness of movement and freezing that are hallmarks of Parkinson's disease.

Until now, such data have been accessible only when a patient's brain is exposed briefly during surgery. But being able to make long-term

ILLUSTRATION BY CHAD HAGEN

neural recordings from human patients may become increasingly important — especially because researchers are experimenting with using DBS as a treatment for many other neurological conditions, including depression, obsessive–compulsive disorder and Tourette's syndrome. The networks involved in such disorders are even less well understood than those involved in Parkinson's disease, says Helen Mayberg, a neurologist at Emory University in Atlanta, Georgia. Devices such as Donobedian's could change that, allowing scientists to start to understand just how unhealthy neural networks misfire in different diseases, and what DBS actually does to the brain. "Every disease will be different and one size won't fit all," Mayberg says. "The new technology is going to enable progress exponentially."

Eventually, adds Bronte-Stewart, engineers could use the new-found knowledge about brain networks to build even more-advanced brain implants — devices that could interpret the neural signals they record, monitor their own effectiveness and generate personalized treatments.

"This is such an exciting time," she says. "This is the first time we're really getting a window into the brain."

> *"This is the first time we're really getting a window into the brain."*

'BLACK BOX' BEGINNINGS

The roots of DBS reach back to the 1960s, when Parkinson's disease was commonly treated with surgery to remove or destroy certain brain regions. To pinpoint which areas to target in each patient, some neurosurgeons began to experiment with electrical stimulation. They discovered that the delivery of rapid pulses to the basal ganglia — a cluster of structures including the STN — could markedly reduce the patient's tremors. By the late 1980s, long-term brain stimulation started to emerge as an alternative treatment to surgery[1]. DBS has since been approved for the treatment of Parkinson's and other movement disorders by both the US Food and Drug Administration (FDA) and European regulators, and has been used in more than 100,000 people.

The biological mechanism underlying DBS remains mysterious, and is a subject of controversy. "We've been guessing a lot over the last decade or two," says Michael Okun, a neuroscientist at the University of Florida in Gainesville. "It would be premature for anyone to claim they know exactly how the therapy works."

There are some clues, however. For example, DBS is not thought to mimic any natural signals in the brain. The high-frequency pulses — delivered at 130–180 times per second for Parkinson's disease — exceed the 1–100-hertz frequency range of most natural neural communications. Furthermore, with each 60–90-microsecond burst, DBS typically delivers several orders of magnitude more current than any neuron or groups of neurons can produce.

And it does not seem to produce permanent changes in the brain, at least not when applied to Parkinson's disease, currently one of the most common targets of the technology. Turning on the current can produce immediate relief from symptoms such as tremor and rigidity. But in many people, symptoms return seconds or minutes after the device is turned off, or the battery runs out — which happens every 3–5 years. Nor does the therapy halt the progressive neurodegeneration associated with the disease; in the long run, patients will typically succumb to symptoms that are not well treated by DBS, such as cognitive deterioration.

From the evidence gleaned so far, researchers suspect that DBS does more than affect neural tissue at the site of the electrodes: it somehow disrupts pathological signals that reverberate through multiple brain regions, corrupting their communications (see 'Circuit training').

That theory meshes with the emerging view that Parkinson's disease, as well as depression and many other neuropsychiatric conditions are best understood as network dysfunctions. "That's a really important realization that has caught on in the last five years," says Cameron McIntyre, a biomedical engineer at Case Western Reserve University in Cleveland, Ohio. Indeed, it has helped to launch two major neuroscience efforts in the past year: the US Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative, and the European Union's Human Brain Project.

The primary target of DBS for Parkinson's disease, for example — the STN — sits in the middle of a highly interconnected brain network that helps an individual to control his or her motions. There is some evidence[2] that as Parkinson's destroys neurons in the basal ganglia, the activity of groups of cells in the STN and across this sensorimotor network becomes abnormally synchronized, locking at certain frequencies. DBS seems to release them from these activity patterns, as do some of the drugs that relieve Parkinson's symptoms[3,4].

Recordings from the new generation of neurostimulators are poised to elucidate these mechanisms, not just for Parkinson's but also — as DBS applications broaden — for psychiatric conditions. The data could help to resolve concerns about the wisdom of expanding the treatment's usage. Although the sensorimotor network involved in Parkinson's disease has been mapped in great detail, says Joseph Fins, a medical ethicist at Weill Cornell Medical College in New York City, much less guidance is available on how best to apply the technology to other disorders. "There has got to be a biological rationale for what you're intending to do," he says.

But others argue that controlled testing of DBS in humans need not wait for complete or near-complete understanding of the relevant networks. "As a clinician, that's not really the important question," says Benjamin Greenberg, a psychiatrist at Brown University in Providence, Rhode Island. "The real questions are: do these treatments help people? Are they safe?"

Okun adds that, unlike the field of movement disorders, the mechanistic study of neuropsychiatric disorders has been slowed by a lack of realistic animal models. "If we're going to move forward with some of these human diseases, we are going to have to use humans — in a very careful way, of course," he says.

ZOOMING IN

Mayberg has been doing just that for more than a decade. In 2005 she published one of the first studies on the use of DBS to alleviate severe, treatment-resistant depression[5]. Since then, she has mainly focused her experiments on a structure known as the subgenual cingulate, in which elevated metabolism has been shown to correlate with the severity of a patient's depression[6]. She estimates that the use of DBS in this region and elsewhere has successfully eased symptoms in 40–60% of the roughly 150 cases of depression reported on so far. But in recent years, her group has begun to do better by using brain imaging to map the dense web of nerve fibres zigzagging through and around the subgenual cingulate, which connects to regions involved in learning, motivation, appetite and sleep. Combining this information with the effects seen in patients, Mayberg is zeroing in on millimetre-scale differences in electrode placement that can make the difference between success or failure.

Potentially, she says, new implants such as the device being tested by Bronte-Stewart could help her team to do even better, allowing researchers to monitor patients' condition in real time and fine-tune the stimulation pulses to maximize benefit. "There may be an optimal tuning frequency for a given person, and it may not be the same for everyone," she says.

Creating personalized DBS treatments is a top priority in this field. Just before Donobedian's meeting with Bronte-Stewart, his neurologist, Camilla Kilbane of Stanford University, spends half an hour tuning the device's stimulation settings to address his symptoms.

Using a short-range radio device, she programs a pulse generator implanted in Donobedian's upper chest. The generator — about half the size of a deck of cards — sends electrical pulses through insulated wires that run under the skin of his neck and scalp, and into his brain. Kilbane has already determined during a previous visit the subset of electrode

contacts she wants to tweak, and Donobedian has stopped taking his supplementary Parkinson's drugs overnight so that Kilbane can cleanly isolate the effects of neurostimulation.

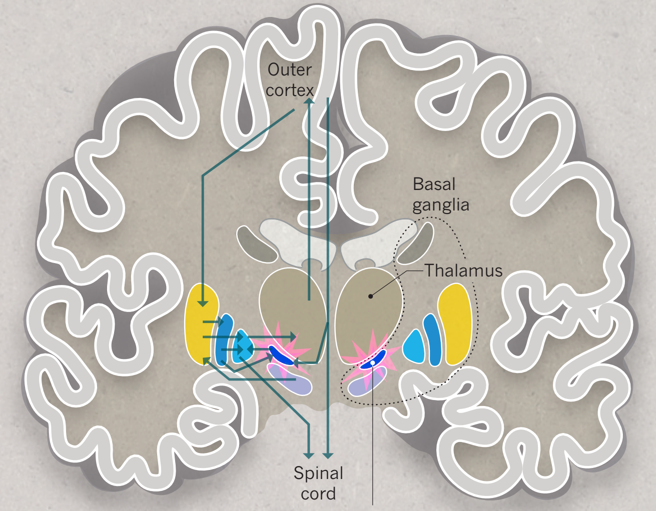
As she drops the voltage and the implant can no longer overcome Donobedian's tremors, his hands and feet begin to quiver again. Within seconds, the tremors grow and spread, until his arms clap against his sides and his shoes tap the linoleum floor. Kilbane clicks the voltage up again, and Donobedian's limbs calm down — but then his arms begin to tingle, a common side effect of DBS. At intermediate voltages, his right leg stops shaking, but the other continues to tremble.

"It's stubborn, that left foot!" remarks Kilbane. She spends another 10 minutes inching the voltage up and down, gradually homing in on an optimal setting. Even after this, Donobedian may need to return in the coming months for further fine-tuning.

"What we have right now for DBS works, but it's very much the first generation," says Bronte-Stewart. She and others are using the new recording-capable DBS implants as a stepping stone towards 'closed-loop' neurostimulators — devices that can continuously track an individual's brain activity and automatically optimize settings as needed in real-time. As a first step, the Stanford group is beginning to mine the electrical recordings downloaded wirelessly from the implants in Donobedian and other patients to find patterns that correlate with different Parkinsonian symptoms. They are also looking to see how these patterns might change in the context of different actions, such as sitting, standing and walking — data that could not be obtained with bulky hospital machines. Indeed, Bronte-Stewart says, there may not be just one set of 'optimal' stimulation parameters. "We may find out there are different frequency ranges that are better for different functions," she says.

CIRCUIT TRAINING

In Parkinson's disease, cell death in the basal ganglia disrupts normal brain signalling, and is thought to send neural static through the large-scale network that controls movement (arrows; only one side shown). Deep brain stimulation seems to repress that static.



SMARTER STIMULATION

As scientists collect more data, some manufacturers are already starting to make strides in closed-loop technology. Last November, the FDA approved the first closed-loop, implantable neurostimulator for intractable epilepsy, another disorder attributable to network dysfunction. The device, made by NeuroPace in Mountain View, California, monitors neural networks for the first sign of abnormal activity — which in some patients originates again and again at one or a few 'epileptic foci' — then responds with a pulse of electrical current to prevent a seizure. "We use stimulation to disrupt that abnormal activity so that it doesn't get picked up by the adjacent neurons," explains Frank Fischer, the company's chief executive.

But Fischer concedes that, whatever the device might do for epilepsy treatment, the technology is not immediately applicable to other conditions. Epilepsy is a comparatively simple disorder, generally consisting of discrete episodes of abnormal brain activity. By contrast, Parkinson's disease involves a mishmash of symptoms that rise, fall and morph over time. Researchers are still searching for the relevant neural signatures in Parkinson's and other diseases, and developing the computational tools required to keep up with changing symptoms.

The first laboratory demonstration of a closed-loop DBS system for Parkinson's disease was reported last year by experimental neurologist Peter Brown at the University of Oxford, UK, for a group of eight patients[7]. Brown plugged the patients' DBS implants into an external machine, which triggered stimulation of the STN only when certain abnormal brain rhythms were detected. This selective stimulation improved the symptoms by almost 30% compared with standard DBS treatments, which stimulate the brain at regular intervals.

"It's far short of being introduced into patients," says Brown of the bulky experimental system, but the demonstration does provide an important proof that the closed-loop concept could work for Parkinson's disease.

In an effort to accelerate the move towards closed-loop technology, the US Defense Advanced Research Projects Agency (DARPA) last October announced a 5-year, US$70-million programme to support the development of novel brain stimulators. As part of the BRAIN Initiative, the project aims to foster brain implants to treat conditions such as post-traumatic stress disorder, anxiety and traumatic brain injury. The agency is looking for implantable devices that can monitor and manipulate neural activity not just at one or a few sites at a time, but across entire functional networks of neurons. Accomplishing this goal will require the development of new types of miniaturized sensor, as well as detailed network models of brain function to interpret data streaming in from multiple brain areas, says DARPA programme manager Justin Sanchez.

Some of those models may eventually grow out of data from researchers such as Kendall Lee, a neurosurgeon at the Mayo Clinic in Rochester, Minnesota. At last year's Society for Neuroscience meeting, he presented a prototype DBS system called Harmoni that can deliver current to one area of the brain while recording electrical and neurochemical responses elsewhere (see *Nature* http://doi.org/rvj; 2013). Because the brain uses both electrical and chemical signals to communicate, explains Kevin Bennet, the lead engineer on the project, monitoring each type of data could provide more complete information about what is going on. The group intends to test Harmoni first in patients with movement disorders. But, ultimately, the scientists hope to extend combined chemical and electrical monitoring to psychiatric disorders. "Those will be the most difficult to treat," says Bennet. "The symptoms are harder to detect and quantify."

Bronte-Stewart projects that testing might begin in about five years for the first implantable, closed-loop DBS devices for Parkinson's disease, with psychiatric applications following close behind. It is not clear whether Donobedian and other current research volunteers could be easily upgraded to those systems; much depends on the precise design of the devices. But even if he does not benefit directly from the data he is generating, Donobedian is glad to participate.

"Somebody had to give to me, to get this far," he says. "If there's a chance for me to give something back without too much effort, I'd like to help." ■ SEE EDITORIAL P.273

Helen Shen *is a freelance writer in Mountain View, California.*

1. Benabid, A. L., Pollak, P., Louveau, A., Henry, S. & de Rougemont, J. *Appl. Neurophysiol.* **50,** 344–346 (1987).
2. Weinberger, M. *et al. J. Neurophysiol.* **96,** 3248–3256 (2006).
3. Whitmer, D. *et al. Front. Hum. Neurosci.* **6,** 155 (2012).
4. Levy, R. *et al. Brain* **125,** 1196–1209 (2002).
5. Mayberg, H. S. *et al. Neuron* **45,** 651–660 (2005).
6. Osuch, E. A. *et al. Biol. Psychiatry* **48,** 1020–1023 (2000).
7. Little, S. *et al. Ann. Neurol.* **74,** 449–457 (2013).

The $1,000 genome

With a unique programme, the US government has managed to drive the cost of genome sequencing down towards a much-anticipated target.

BY ERIKA CHECK HAYDEN

In Silicon Valley, Moore's law seems to stand on equal footing with the natural laws codified by Isaac Newton. Intel co-founder Gordon Moore's iconic observation that computing power tends to double — and that its price therefore halves — every 2 years has held true for nearly 50 years with only minor revision. But as an exemplar of rapid change, it is the target of playful abuse from genome researchers.

In dozens of presentations over the past few years, scientists have compared the slope of Moore's law with the swiftly dropping costs of DNA sequencing. For a while they kept pace, but since about 2007, it has not even been close. The price of sequencing an average human genome has plummeted from about US$10 million to a few thousand dollars in just six years. That does not just outpace Moore's law — it makes the once-powerful predictor of unbridled progress look downright sedate. And just as the easy availability of personal computers changed the world, the breakneck pace of genome-technology development has revolutionized bioscience research. It is also set to cause seismic shifts in medicine.

In the eyes of many, a fair share of the credit for this success goes to a grant scheme run by the US National Human Genome Research Institute (NHGRI). Officially called the Advanced Sequencing Technology awards, it is known more widely as the $1,000 and $100,000 genome programmes. Started in 2004, the scheme has awarded grants to 97 groups of academic and industrial scientists, including some at every major sequencing company.

It has encouraged mobility and cooperation among technologists, and helped to launch dozens of competing companies, staving off the stagnation that many feared would take hold after the Human Genome Project wrapped up in 2003. "The major companies in the space have really changed the way people do sequencing, and it all started with the NHGRI funding," says Gina Costa, who has worked for five influential companies and is now a vice-president at Cypher Genomics, a genome-interpretation firm in San Diego, California.

A GIANT'S LEGACY

The $1,000 genome programme, now close to achieving its goal, will award its final grants this year. As technology enthusiasts look to future challenges, the coming milestone raises questions about how the roughly $230-million government programme managed to achieve such success, and whether its winning formula can be applied elsewhere. It benefited from fortuitous timing and the lack of an entrenched industry. But Jeffery Schloss, director of the division of genome sciences at the NHGRI in Bethesda, Maryland, who has run the programme from its inception, says that its achievements also suggest that there are ways to navigate public–private partnerships successfully. "One of our challenges is to figure out what is the right role for the government; to not get in the way, but feed the pipeline of private-sector technology development," he says.

The quest to sequence the first human genome was a massive

undertaking. Between 1990 and the publication of a working draft in 2001, more than 200 scientists joined forces in a $3-billion effort to read the roughly 3 billion bases of DNA that comprise our genetic material (International Human Genome Sequencing Consortium *Nature* **409,** 860–921; 2001). It was a grand but sobering success. The project's advocates had said that it would reveal 'life's instruction book', but in fact it did not make it possible to interpret how the instructions encoded in DNA were transformed into biology. Understanding how DNA actually influences health and disease would require studying examples of the links between genes and biology in thousands, perhaps millions, more people.

The dominant technology at the time was Sanger sequencing, an inherently slow, labour-intensive process that works by making copies of the DNA to be sequenced that include chemically modified and fluorescently tagged versions of the molecule's building blocks. One company, Applied Biosystems in Foster City, California, provided the vast majority of the sequencers to a limited number of customers — generally, large government-funded laboratories — and there was little incentive for it to reinvent its core technology.

Still, researchers had seen some advances, including robots that replaced some human work and improvements in devices capable of handling small amounts of liquid. At a 2002 meeting convened by the NHGRI, scientists predicted that such developments would drive costs down at least 100-fold over the next five years. But that was not enough.

They debated what price target would make human genome sequencing routine, the kind of thing a physician might order to help diagnose a patient — on a par with a magnetic resonance imaging scan. "Somebody threw out, to great rolling of eyes, 'a thousand dollars'," recalls Schloss.

That seemed too ambitious, given the state of the technology. "The risk associated with that is not one that your normal investor is willing to spend any money on," says Eric Eisenstadt, a retired official from the US government's Defense Advanced Research Project Agency who is now a consultant in Reston, Virginia.

So Schloss and the NHGRI stepped in and began to fund basic research on entirely new methods of sequencing, as well as industrial research to develop these technologies for commercial use. The mixture of applied and academic research within a single programme was uncommon at the National Institutes of Health (NIH), the NHGRI's parent agency. The project was also more nimble than the typical NIH grant programme because it allowed the agency to make small awards for work considered promising but risky. "That flexibility is unusual for the NIH," says Schloss.

Furthermore, the programme provided support to sequencing companies that could compete with Applied Biosystems. One of the companies funded in the first round of grants, 454 Life Sciences of Branford, Connecticut, was the brainchild of entrepreneur Jonathan Rothberg. It aimed to develop a method that was faster and cheaper than Sanger sequencing by using a much simpler sample-preparation procedure and running many sequencing reactions simultaneously on a solid surface. But as he tried to round up funding, Rothberg heard the same refrain over and over from investors. "People said, 'Why would you want to sequence DNA fast? We've already done the Human Genome Project.'"

A $7-million award from the NHGRI allowed the company to commercialize a technology called pyrosequencing, which was the first to begin chipping away at Applied Biosystems' monopoly.

The funding commitments also ultimately helped to convince private investors to enter the market. Stephen Turner, founder and chief technology officer of Pacific Biosciences in Menlo Park, California, says that his company's 2005 NHGRI grant of $6.6 million helped to attract subsequent venture-capital funding. The NHGRI's imprimatur had convinced investors to provide the much more substantial money that the company needed to commercialize its technology, which observes DNA synthesis as it occurs in real time.

"Having experts in sequencing technology give us a favourable score was hugely influential," says Turner.

The NHGRI's investments — typically a few million dollars or less — could not by themselves nurse a technology from lab to market. But they could fund parts of it, such as work on improving a dye, a piece of circuitry or a laser, or tests of combinations of components.

The programme has invested $88 million in technologies based on nanopores and nanogaps. The form of this technology closest to the market involves reading bases as they are threaded through a pore (see *Nature* **456,** 23–25; 2008), a method that has long promised to save costs and time by reading DNA while it is processed. It would negate the need for expensive and slow reactions to make lots of copies of the molecule. But solving basic issues, including how to move the DNA through the pore slowly enough, has been a major challenge. The NHGRI has funded work to overcome these hurdles — including $9.3 million given to collaborators of the company now ushering the concept to market, UK-based Oxford Nanopore Technologies (see *Nature* http://doi.org/rvm; 2014). Turner says that such investments have helped to cut sequencing costs before the technology hits the shelves.

"There's a tremendous amount of altruistic sharing of knowledge."

The $1,000 genome project seeded so many companies and labs that it populated the entire industry with expertise, say sequencing researchers. One of the beneficiaries of that is Illumina in San Diego, currently the market leader in sequencing machines. Illumina, whose technology reads out many short stretches of DNA, has acquired multiple companies and many scientists who were once supported by the NHGRI. "It's through acquisitions that Illumina has become stronger and stronger," says Mostafa Ronaghi, the company's chief technology officer.

But Schloss's programme also forced competitors to exchange expertise at an annual progress meeting that has become a must-attend event. "That meeting is one of the most important venues for keeping an eye on what's happening in sequencing technology development," says Turner. "There's a tremendous amount of altruistic sharing of knowledge that occurs."

CUT-PRICE CAVEATS

Some scientists question certain choices made by the programme. "There's been a lot of money given to the nanopore space, but the objective of nanopore really hasn't hit the mark," says Costa, for example.

Kevin McKernan, who worked with Costa to develop SOLiD — a sequencing technology based on an enzyme that joins pieces of DNA together — points out that many of the companies funded by the $1,000 genome programme ultimately failed. "Their hit rate probably isn't much better than a venture capitalist," he says.

But others give Schloss and the programme credit for spreading their investments over an array of academic and industrial work that is diverse enough to allow progress in the face of failures. Many NHGRI-funded firms are now defunct — including 454, and Helicos BioSciences of Cambridge, Massachusetts — but other grant recipients have moved the field forward, often using ideas generated by the shuttered firms.

"The NHGRI funded smaller companies and academic groups to create a pipeline of technologies," says Ronaghi. "They didn't decide which technologies to bet on."

Working out what the $1,000 genome programme got right has emerged as a key question as Schloss and the NHGRI shape its successor. Sequencing still needs much improvement, especially in terms of quality. For all of Sanger sequencing's high cost, it remains the benchmark for accuracy. And sequencing costs are no longer dropping as quickly as they were a few years ago.

But researchers are optimistic that another technology will emerge to challenge Illumina. Most think, in fact, that the crucial questions for the field will shift away from technology. Now that sequencing is cheap enough to talk about scanning every patient's genome, or at least the protein-coding portion of it, it is still not clear how that information will translate into improved care (see *Nature* http://doi.org/rvq; 2014). These more complex issues will require another great leap in genomic science — one that could make the trouncing of Moore's law seem easy. ■ SEE EDITORIAL P.273

Erika Check Hayden *writes for* Nature *from San Francisco, California.*

ENERGY Governments must keep backing the renewables boom **p.297**

HOMININS Was the hunt for water the key driver in human evolution? **p.303**

VISUALIZATION Exhibition celebrates history of scientific graphics **p.304**

REPRODUCIBILITY Hypothesis-driven projects risk data discounting **p.306**

ILLUSTRATION BY DAVID PARKINS

Industry–funded academic inventions boost innovation

Brian D. Wright and colleagues present data challenging the assumption that corporate-funded academic research is less accessible and useful to others.

Governments have long encouraged university–industry collaboration, hoping to spur innovations that bring jobs, investment and life-enhancing products[1]. At the same time, shrinking government budgets for science have forced universities to look to other sources of funding. According to the US National Science Foundation, in 2012, industry supplied just over 5% (some US$3.2 billion) of US research universities' annual expenditure[2].

But the role of corporations in academic research is controversial. For example, when oil company BP announced in 2007 that it would pay $500 million to fund a decade of alternative-energy research by a consortium headed by the University of California, Berkeley, this prompted a backlash. Fearing that industry money would contaminate the public institution's research agenda, many students, staff and members of the community picketed the campus with a 2.5-metre Trojan horse. An earlier agreement between the department of plant and microbial biology at Berkeley and the Swiss pharmaceutical firm Novartis sparked similar opposition. At the 1999 graduation ceremony, about 100 students displayed the company's logo on their mortarboards, protesting that the department had been bought by corporate interests.

There are reasons to be cautious about corporate sponsorship of academic research[3]. The tobacco, food, pharmaceutical and other industries have been shown to manipulate research questions and public discourse for their own benefit and even to suppress unfavourable research[4]. And companies may shift university researchers towards narrow corporate interests. If the results of research are privately held, others cannot exploit them.

Conversely, some feel that overly restrictive university technology-transfer policies stifle productive deal-making between firms and academic researchers[5]. Some advocate that a university's intellectual property should ▶

be managed by an outside agency[6], or else handed over directly to researchers or to the companies funding their work[7].

Data to inform this debate are hard to come by. Individual universities may track patents and licences at their own institutions, but these data sets are generally small and confidential. The prevailing assumption is that corporate-sponsored inventions and the information associated with them are less accessible and less useful to others than inventions sponsored by the government or non-profit organizations.

Here we offer empirical evidence to the contrary. Our analysis suggests that corporate-sponsored research is surprisingly valuable for further innovation. Data collected over 20 years at nine campuses and three national laboratories administered by the University of California show that corporate-sponsored inventions are licensed and cited more often than federally sponsored ones.

Although results might differ at other academic institutions, these findings should allay concerns that corporate sponsorship turns leading universities into corporate vassals. Collecting and combining data from a larger sample of institutions could help to both explore what corporations hope to gain from funding academic work, and suggest how universities can best manage research sponsorships.

TECHNOLOGY TRANSFER

Like most universities, the University of California requires faculty members and other researchers to disclose any invention that has commercial potential to one of its offices of technology transfer (OTTs), and to list funding sources for the project that led to it. Under these terms, an invention is anything that a researcher feels could be patented or is otherwise valuable as intellectual property: it might be a material, a method, or an animal or plant. The OTT then determines whether to pursue intellectual property protection on the university's behalf and negotiates contracts with potential licensees.

From 1990 to 2005, University of California faculty members, staff and students, and employees of the three associated national laboratories disclosed 12,516 inventions to their OTTs. Of these, nearly 1,500 were supported, at least in part, by corporate funds. Under strict terms of confidentiality, the central OTT provided us with data on these disclosures, and on related licensing activities, until the end of 2010. From 1990 to 2010 the University of California campuses accounted for up to 9% of total US academic research expenditure. Collectively, they obtained more issued patents than any other US academic institution. In lists compiled annually by the US Patent and Trademark Office, the multi-campus University of California system often had more than twice as many patents as the second-largest patent producer in academia (generally the Massachusetts Institute of Technology in Cambridge).

Of all inventions generated at the University of California, 20% are linked to at least one licence, and nearly 25% were eventually patented. Inventions with no sponsor information were the least likely to yield either licences (13%) or patents (17%). We believe that most of these inventions came either without extramural support or with federal support, which is such a common situation that inventors or technology-transfer agents may not note it explicitly. Corporate-sponsored inventions resulted in licences (29%) and patents (35%) more frequently than federally sponsored ones (22% and 26%, respectively). The rates are higher still for inventions with both types of sponsor; 36% were licensed and 43% patented (see 'Licensed and cited'). Results were similar across technical fields. More than two-thirds of classified technologies relate to biological, pharmaceutical and chemical advances, a distribution that is consistent with other leading research universities (for the complete results see Supplementary information; go.nature.com/o99eua).

Although corporate-sponsored inventions are more likely to be patented, that does not mean that corporate support makes inventions more patentable. Instead, corporations might select projects that are more likely to produce patentable inventions.

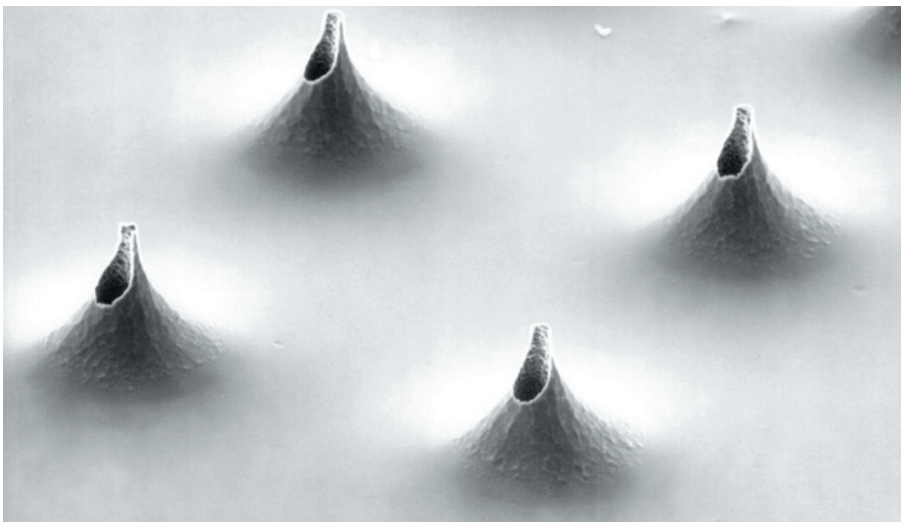
Corporations typically get priority to negotiate licences to the inventions they sponsor, and 86% of the licences to the sponsors are exclusive, meaning that the university agrees not to grant the same rights to multiple licensees. Of licensed inventions associated with some form of intellectual property, 78% were licensed exclusively, consistent with the share of 79% reported for licensing of patents funded by the National Institutes of Health[8].

Nevertheless, our analysis did not support our original assumptions that licences to industry-sponsored inventions would be likely to be exclusive, or that sponsors would snap up the lion's share of exclusive licences. First, the overall percentage of corporate-sponsored inventions licensed exclusively (74%) is not higher than for those with solely public funding (76%). Second, half of the exclusive licences for corporate-sponsored inventions seem to be to third parties (although we cannot be sure that we identified all the sponsor-controlled firms in the data). Apparently, even the inventions that sponsors leave on the table have substantial value, because these licensees usually bear significant costs of patenting, plus agreements to pay future royalties.

Another surprise is that corporate-sponsored inventions spur more 'knowledge spillovers', on average, than federally sponsored research, according to forward citation rates, the most widely used metric for patent quality and value. Forward citations show how many times one patent is cited in subsequent patents. Each corporate-sponsored invention generated, on average, 12.8 forward citations if licensed to a third party (more if licensed by the sponsor), compared with 5.6 for federally sponsored inventions. This runs counter to the expectation that corporate-sponsored inventions have narrow applications, and so create more private benefits but few benefits for others.

USING UNIVERSITIES

This analysis does not address how corporate funds affect universities' research agendas, but it does dispute the idea that corporations tie up all sponsored inventions to restrict access. Instead, high patent citation rates for corporate-sponsored inventions suggest that firms are funding exploratory research. Work by sociologist James Evans



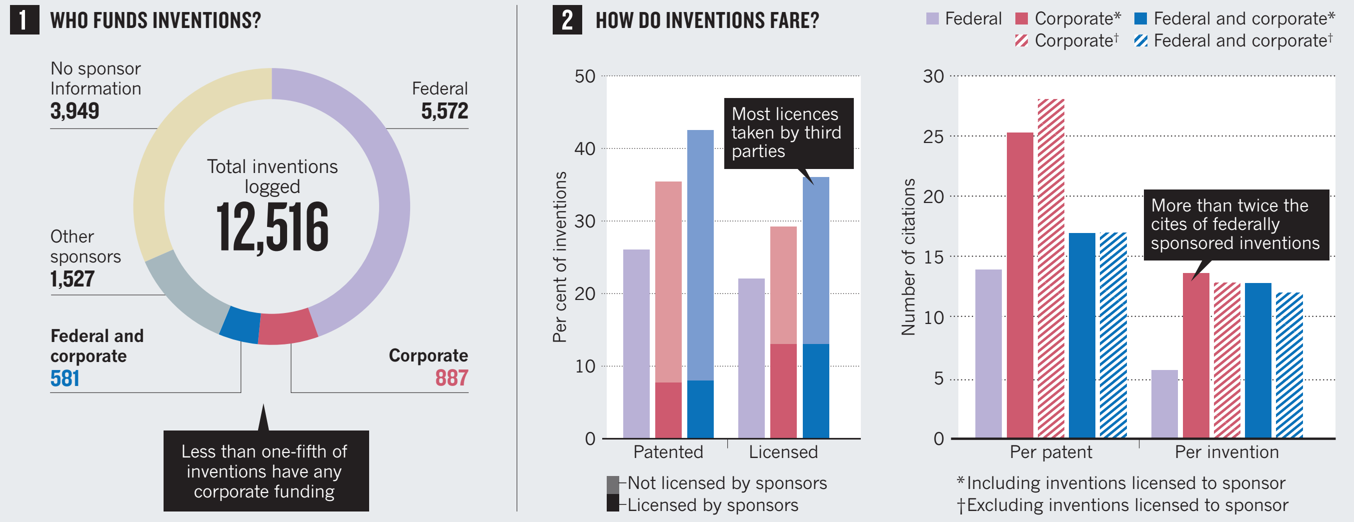
STOEBER, B. & LIEPMANN, D. *J. MICROELECTROMECH. SYST.* **14**, 472–479 (2005)/IEEE

Microneedle fabrication, the subject of one of the most highly cited University of California patents.

UC PATENT TRACKING SYSTEM DATABASE

LICENSED AND CITED

Of the 12,516 inventions logged by technology-transfer offices of the University of California system between 1990 and 2005, inventions with only federal funding were less likely to be patented or licensed than those with corporate or corporate and federal funding, and had lower patent citation rates.



at the University of Chicago in Illinois suggests that corporations turn to universities to investigate areas outside their core strengths, investing in speculative science in the hope of finding profit opportunities[9].

In fact, Evans argues that corporations actually urge academics to explore further afield than they might otherwise. Although academics may act conservatively to gain acceptance of peers, papers and grant proposals, he writes[9], "industry partnerships draw high-status academics away from confirming established theories and towards speculation".

For example, the $500-million research grant from BP to the Berkeley-led consortium was intended to explore biofuels from cellulose in plants or crop residues, an area in which BP had virtually no expertise. In such cases, many resulting inventions might turn out to be informative to other researchers, but irrelevant to the firm's business strategy.

In such cases, other firms' subsequent work on an invention can be more valuable to sponsors than exclusive access. For example, preliminary work by Yongdong Liu, a PhD candidate at Berkeley, suggests that information-technology company IBM discloses innovations on the periphery of its expertise without patenting them, but often cites non-IBM patents building on the disclosed innovations. Similarly, some major drug companies contributed to the publicly funded Human Genome Project, reasoning that faster access to results would accelerate its ability to develop drugs, even if those results were openly available.

Acquiring intellectual property is not necessarily the prime focus of corporate sponsors. Companies also value sustained relationships with leading scientists and associated opportunities to identify and recruit talented employees. The University of California–Novartis agreement apparently generated no licences for the company, and Novartis representatives reportedly did not exert any apparent influence on the selection of projects it funded[10].

Joint federal–corporate sponsorship may stem from more-focused goals. We understand that they often arise from projects initiated by federal funding agencies, with corporate sponsors recruited to develop early, promising work into practical applications. For example, if a federally sponsored gene-screening programme finds an attractive drug target, corporations might support projects to screen drug candidates against that target. This kind of focus would explain why inventions in this category are the most likely to be licensed (even by third parties) but not more highly cited.

The large share of third-party licences suggests that the University of California successfully markets inventions and also negotiates agreements to keep corporations from locking them up unduly. This task is probably facilitated by the fact that many sponsoring firms seem to recognize that sharing exploratory research can be in their own interests.

To assess whether these findings generalize to other academic institutions, data from other research universities are needed. We advocate a project to pool similar data from a large sample of other research universities, with solid confidentiality safeguards, for empirical analysis. Such work could evaluate whether, for instance, groups of smaller or less research-oriented institutions would be better served by outsourcing to a single technology-transfer institution.

Universities setting up contracts with corporations need to be vigilant in their mission to generate and transfer knowledge, but they should not assume that companies are focused mainly on tying up intellectual property. Those that do will miss fruitful opportunities for collaboration with firms willing to fund projects from which many others will probably benefit. ■

Brian D. Wright *is professor of agricultural and resource economics at the University of California, Berkeley, USA.* **Kyriakos Drivas** *is a postdoctoral research economist at the Agricultural University of Athens, Greece, and a research fellow at the University of Piraeus, Greece.* **Zhen Lei** *is assistant professor of energy and environmental economics at the Pennsylvania State University in University Park, USA.* **Stephen A. Merrill** *directs the US National Academy of Sciences' Program on Science, Technology, and Economic Policy in Washington DC, USA.*
e-mail: bwright@berkeley.edu

1. President's Council of Advisors on Science and Technology *University–Private Sector Research Partnerships in the Innovation Ecosystem* (OSTP, 2008); available at http://go.nature.com/hilyum.
2. National Science Board *Science and Engineering Indicators 2014* 5–13 (NSF, 2014).
3. Washburn, J. *University, Inc: The Corporate Corruption of Higher Education* (Basic Books, 2005).
4. White, J. & Bero, L. A. *Stanford Law Policy Rev.* **21,** 105–133 (2010).
5. Kramer, D. *Physics Today* **61,** 20–22 (2008).
6. Litan, R., Mitchell, L. & Reedy, E. J. *Innov. Policy Econ.* **8,** 31–57 (2008).
7. Foley, H. C. *Res. Technol. Mgmt* **55,** 12–17 (2012).
8. Pressman, L. *et al. Nature Biotechnol.* **24,** 31–39 (2006).
9. Evans, J. *Am. J. Sociol.* **116,** 389–452 (2010).
10. Busch, L. *et al. External Review of the Collaborative Research Agreement between Novartis Agricultural Discovery Institute, Inc. and The Regents of the University of California* (Inst. Food and Agricultural Standards, Michigan State Univ., 2004); available at http://go.nature.com/sgoc1a.

COMMENT

A solar power plant on a former military air base in Finowfurt, Germany.

Back the renewables boom

Low-carbon technologies are getting better and cheaper each year, but continued public-policy support is needed to sustain progress, says **Jessika E. Trancik**.

A new battery is rarely greeted with as much excitement as the latest smartphone or a new drug. The energy industry is widely perceived as sluggish, a provider of basic services and lacking creativity. In fact, a brighter reality is emerging — government support for energy-technology development is paying off.

Public policies to encourage the development and adoption of renewable-energy technologies are essential, because low-carbon performance is not visible to most consumers and carbon is not priced in the global market. Yet there is a widespread lack of confidence in public-sector efforts to spur innovation, as a result of the mixed record of governments in picking winners and losers among technologies[1].

Some governments are considering reducing their support for renewable-energy projects. The future of the US tax credit for new wind energy is uncertain; the United Kingdom is debating scaling down subsidies for some renewables and relaxing its targets for carbon-emissions reductions, and Spain has abandoned its incentives programme and electricity-price commitments for renewable-energy power plants. The countries of the European Union disagree on a common binding target for the adoption of renewable energy by 2030.

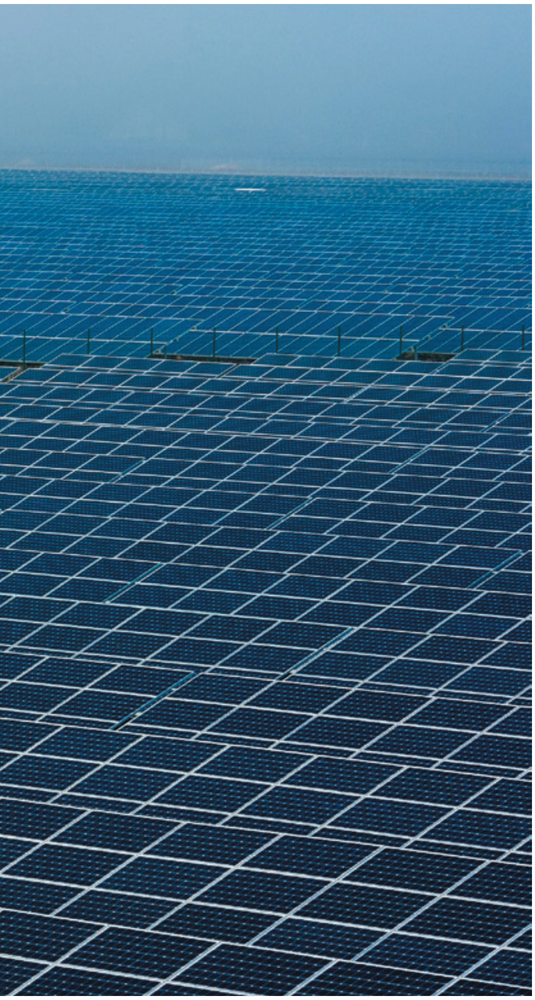
But now is not the time to cut government support for renewables. Each day that we delay implementing low-carbon energy technologies we increase the likelihood of damage from climate change — from storms and floods to forest fires, .

The response of the global energy industry to even modest policy interventions has been remarkable. Led by China, Europe, the United States and Japan, the alternative-energy sector is booming worldwide[2–4]. Solar and wind technologies have improved most rapidly in the past three decades, with photovoltaics a hundred times cheaper today than in 1975.

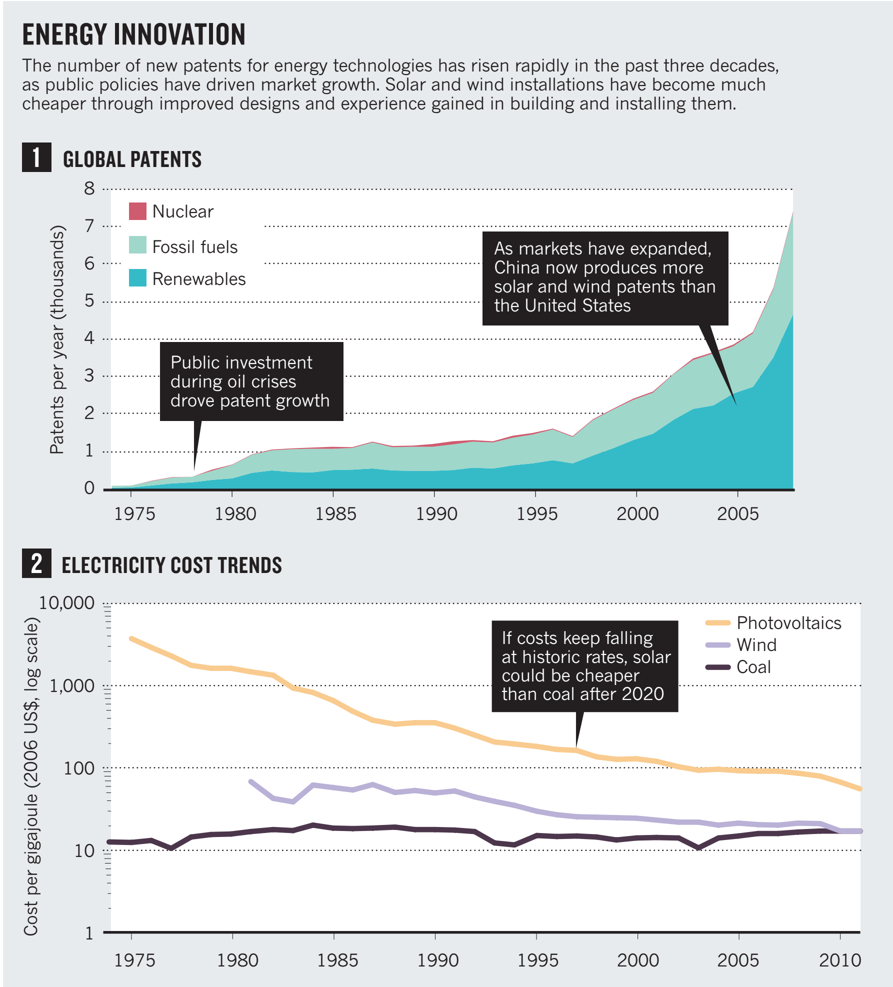
Governments should help to maintain this progress. Research funds and policies to boost markets will mature new energy industries and promote the next generation of low-carbon technologies.

RAPID INNOVATION

The speed of energy-technology innovation is only just coming to light as long-term data sets become available. My analyses of 30 or more years of data[2–4] show that the costs of



DPA PICTURE ALLIANCE ARCHIVE/ALAMY



SOURCE: REFS 2–4

renewable-energy technologies have fallen steeply. Photovoltaic module costs have plunged by about 10% per year over the past 30 years and the costs of wind turbines have fallen by roughly 5% per year. Production levels for both technologies have risen by about 30% per year on average.

The technical advances responsible have been driven by public policies and industry's responses to them. Governments spend a relatively modest amount on renewable-energy research, roughly US$5 billion per year globally, which is less than one-tenth the amount allocated to health research. But government incentives are essential for market growth; they drive private-sector investments in clean-energy technologies of about $250 billion per year globally.

Despite this success, lawmakers in many countries are questioning public support for clean energy. Some in the United States are urging that such support should be limited to funding basic research and development in universities and government labs. They cite the recent failure of a few prominent energy companies, such as Solyndra, which received government grants or loans in their early days. Critics forget that game-changing technologies are high-risk ventures; some failures are inevitable.

Some technologies are more open to improvement than others[2–7]. Compact, modular systems, such as photovoltaics and electronics, are easily experimented on[5,7]. And processes that may be achieved through alternative designs or materials offer more avenues for advancement. The diversity of semiconductors, for example, is behind the recent development of high-efficiency perovskite solar cells. Other technologies are harder to improve. Those with high commodity costs, such as coal-fired electricity[6], soon hit cost floors in the marketplace.

"Game-changing technologies are high-risk ventures; some failures are inevitable."

Photovoltaic systems and wind turbines are therefore better candidates for sustained cost reduction than large nuclear or coal plants[7]. The lower price of solar cells today is due to increasing the efficiency with which sunlight is converted to electricity within modules, less manufacturing waste and greater economies of scale.

Wind turbines have seen similar progress, reaching higher wind speeds at greater heights to deliver more energy per cost of installed unit. Indeed, wind energy now competes economically with fossil-fuelled thermal power plants in several places, including Texas, and in Denmark it supplies 30% of electricity consumption.

Knowledge about how to design, build and integrate these technologies into the energy infrastructure has also grown. For example, it takes roughly half the time to install a solar system in Germany compared with the United States, thanks to a more experienced workforce and streamlined permit processes[8].

The recent growth in energy patents reflects the increasing knowledge generated in laboratories and on the manufacturing floor. The numbers of patents for solar and wind technologies have risen globally by about 15% and 20% per year on average, respectively, in the past decade[2] — rates comparable with information technologies.

A link between public investment and patent filing is evident in historic records[2]. In the 1970s and early 1980s, patent numbers

for solar and wind technologies rose by 20% and 15% per year on average globally as public money in Japan, the United States and Europe was pumped in to find alternatives to oil during the energy crises. When global energy-research funding dropped by 70% between 1980 and 1990, patenting rates levelled off, causing concern that the development of renewables would slow[9].

Patenting rates remained flat for nearly two decades until the early 2000s, when they rose again[2] (see 'Energy innovation'). This time the reason was rising investment from industry, in response to various government incentives and regulations.

Certain nations are now clear leaders in innovation. Japan dominates in terms of cumulative energy patents filed. But in the past ten years' growth has been driven by China and the United States, which together account for roughly 60% of renewables patents published globally per year and 60% of all energy patents. China has been the front runner in coal patents for 15 years and, in the past decade, has overtaken the United States in annual patent numbers for wind and solar as well. Europe has seen a tripling of renewables technologies patents in the past decade, whereas the region's patents related to fossil fuels have declined.

Underlying these cost and patenting trends is a diversity of national or regional policies[10], including research funding, market incentives such as feed-in tariffs, subsidies and regulated adoption levels. China has set a target to supply 15% of energy for electricity production with renewables by 2020, and Germany plans to produce 35% of its electricity through renewables by 2020. In the United States, 29 states and the District of Columbia have adopted renewables portfolio standards that require specified installation levels of renewable energy.

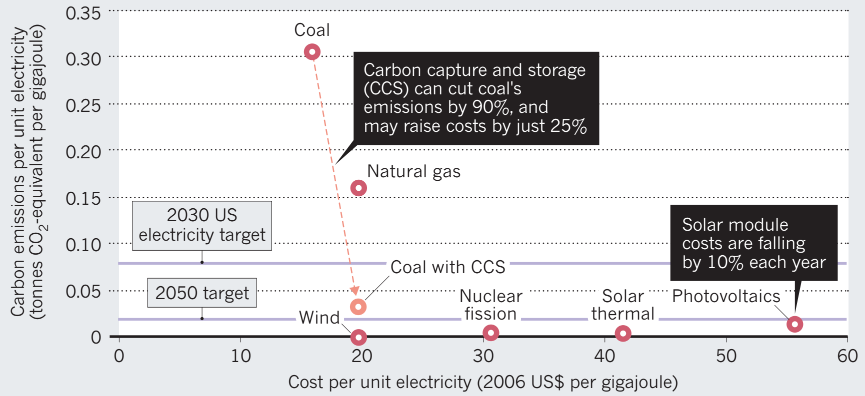
MAINTAIN MOMENTUM

Even with extreme energy-efficiency measures, developed countries will need to meet 75–100% of their power demand with carbon-free sources by 2050 (ref. 4), compared with 30% globally today, to hit emissions targets. The remainder could be produced with clean coal (with carbon capture and storage) and natural gas.

Sustained public policies and monitoring of technology performance are needed to support further progress in a variety of low-carbon technologies for electricity, transport and heating until they can support themselves. Wind and solar energy have delivered the greatest bang per buck so far, in terms of improvements in cost, patenting and market growth, and further developments should be supported. Because both types of power are intermittent, storage technologies should also be priorities for government investment.

EVALUATING PERFORMANCE

The performance of different energy technologies in terms of average emissions and cost intensities can be compared against carbon-intensity targets (here for the United States).



SOURCE: REF. 2

The rapidly changing landscape of energy-technology patents and manufacturing costs, exemplified by the rise of China in these areas, cautions nations and corporations to build experience quickly in these fast-moving markets. Such momentum is evident in the highly efficient assembly lines now running at Yingli Solar's headquarters in Baoding, China, and in petrol stations in Sweden that offer drivers a choice of natural gas, biofuels, diesel and petrol. Students at my university are pitching clean-energy ideas to start-up investors with a sophistication that was rare even among experts ten years ago.

Ways to measure progress towards low-carbon goals will be needed so that analysts and policy-makers can identify top-performing technologies and policies. Technological improvements can be quantified by the use of a 'cost-carbon curve' (see 'Evaluating performance'). In this way, the carbon (or water or land) intensity of technologies can be compared with their competitiveness, and evaluated against performance targets that depend on particular global or regional concerns such as emissions (or water consumption or land use)[4].

Wind power is on the rise in China.

IMAGINECHINA/REX

Any major energy transformation will involve stumbles — from technologies that flop or companies that close to imperfect policy instruments, such as the teething troubles with the European emissions-trading scheme. Encouraging a diversity of innovations and monitoring progress will lessen these risks.

Governments should recognize the boom in energy-technology development and continue to support it. As history suggests, the resulting momentum could be enough to largely decarbonize the world's energy supply by mid-century. ■

Jessika E. Trancik *is an assistant professor in the Engineering Systems Division, Massachusetts Institute of Technology, Cambridge, Massachusetts, and at the Santa Fe Institute, New Mexico, USA.*
e-mail: trancik@mit.edu

1. Fri, R. W. *Energy J.* **24,** 51–74 (2003).
2. Bettencourt, L. M. A., Trancik, J. E. & Kaur, J. *PLoS ONE* **8,** e67864 (2013).
3. Nagy, B., Farmer, J. D., Bui, Q. M. & Trancik, J. E. *PLoS ONE* **8,** e52669 (2013).
4. Trancik, J. E. & Cross-Call, D. *Environ. Sci. Technol.* **47,** 6673–6680 (2013).
5. McNerney, J., Farmer, J. D., Redner, S. & Trancik, J. E. *Proc. Natl Acad. Sci. USA* **108,** 9008–9013 (2011).
6. McNerney, J., Farmer, J. D. & Trancik, J. E. *Energy Policy* **39,** 3042–3054 (2011).
7. Trancik, J. E. *Environ. Res. Lett.* **1,** 014009 (2006).
8. Seel, J., Barbose, G. & Wiser, R. *Why Are Residential PV Prices in Germany So Much Lower Than in the United States? A Scoping Analysis* (Lawrence Berkeley National Laboratory, 2013).
9. Margolis, R. M. & Kammen, D. M. *Science* **285,** 690–692 (1999).
10. Trancik, J. E., Chang, M. T., Karapataki, C. & Stokes, L. C. *Environ. Sci. Technol.* **48,** 27–35 (2014).

MAURICIO ANTON/SPL

An early human relative, *Homo antecessor*, takes advantage of game animals' attraction to water (artist's impression).

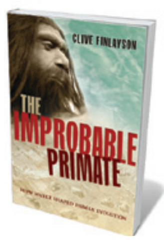
HUMAN EVOLUTION

Just add water

Richard G. Roberts is intrigued by the idea that early humans gained an edge by 'hunting' for lakes and rivers.

We humans are a thirsty lot. Our bodies are between 55% and 60% water, and we need to drink 2 to 3 litres of fluids per day to remain healthily hydrated. In the prehistoric past, our ancestors sourced their supplies from lakes, wetlands and rivers, but these amount to just 0.01% of all water on Earth today. Fresh water has always been scarce. In *The Improbable Primate*, Clive Finlayson argues that the progressive desiccation of the planet over the past few million years — especially three shifts around 2.8 million, 1.8 million and 0.8 million years ago — was the driving force behind our emergence. And, he posits, it triggered the extinction of our closest evolutionary cousins, the Neanderthals, about 30,000 years ago.

The field of human evolution abounds with schemes for how we came to be the dominant primate. When and why did we leave tropical rainforests and take our first steps onto the savannah? How did we acquire our competitive edge over other apes and hominins (primates more closely related to humans than to chimpanzees), enabling us to disperse around the globe and occupy almost every conceivable habitat? The evolution of our large brains, bipedal gait and dexterity with tools have been attributed to all manner of causes — most often some change in the external environment, but also transformations in cultural practices, such as increased meat consumption or cooking, and population dynamics.



The Improbable Primate: How Water Shaped Human Evolution
CLIVE FINLAYSON
Oxford University Press: 2014.

With his "Water Optimization Hypothesis", Finlayson — a zoologist by training — pins his flag securely to the environmental mast. He blazes a trail through the past 16 million years of human ancestry, from fruit-eating apes roaming the rainforest canopy to Australia's desert-dwellers, exemplars of human survival on the driest inhabited continent and "the crowning achievement of almost 2 million years of evolution". Primates are experts at locating key resources in patchy environments, from ripe fruit in the rainforest to freshwater oases on the open landscape. Natural selection rewards those that can exploit these patches most efficiently. Thus were born the "rain chasers": humans built for endurance running and long-distance walking, equipped with portable and perishable objects to suit their nomadic lifestyle. The resulting increased mobility and large-scale consumption of meat by our ancestors — perhaps due to animal-hunting or scavenging of carcasses concentrated around watering holes — are unrivalled among primates. Hence the tag 'improbable'.

Was water the evolutionary driver or is it merely one of the key habitat components that we require to survive? Answers cannot be sought solely in hominin fossils, which are exceptionally rare. So Finlayson scours the archaeological and palaeontological literature for common features of what our ancient forebears called home: some combination of woody cover, open spaces and fresh water. In 1925, the pioneering palaeoanthropologist Raymond Dart proposed that these habitat elements were essential to early human evolution in southern Africa. Finlayson extends this list to the needs of later hominin lineages, adding rocky outcrops, which proved particularly ▸

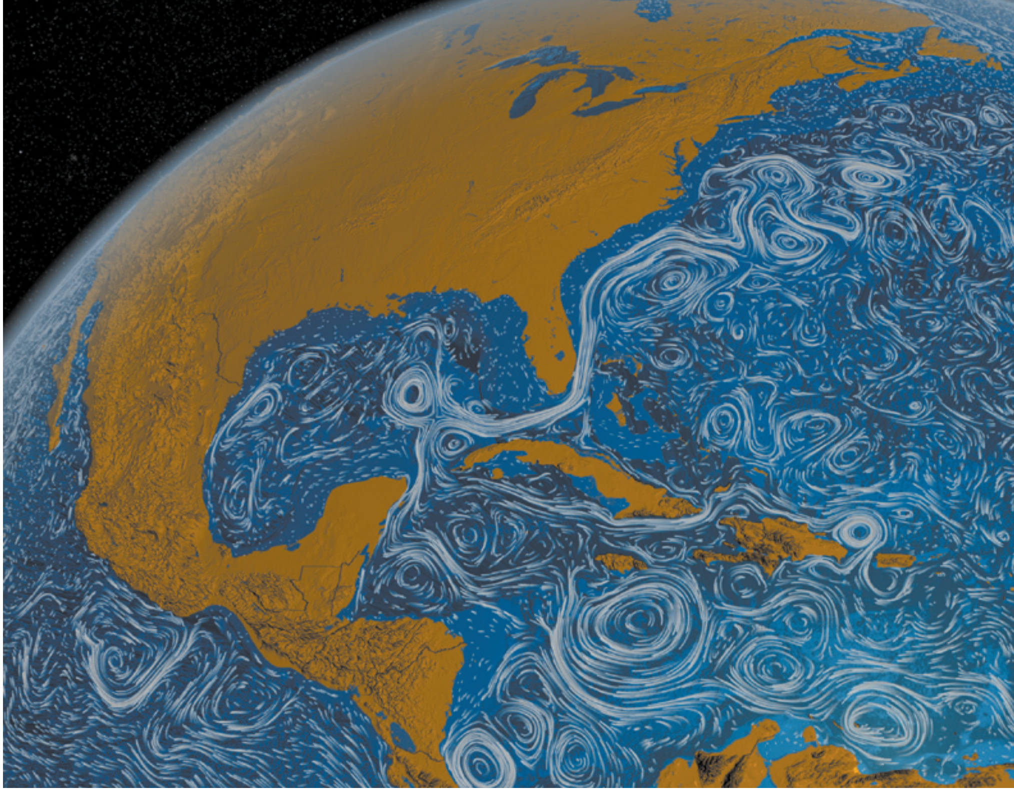
popular with cool-climate Neanderthals occupying the mountain ranges of Eurasia.

Certain regions loom large in *The Improbable Primate* as launch pads for nature's experiments in human evolution: northeast Africa and Arabia (dubbed "Middle Earth", but free of Hobbits), the mountain chains of mainland Asia and the continental shelves of southeast Asia, exposed when sea levels fell during each ice age. In this largely descriptive account of human biogeography, Finlayson paints with a broad brush. Defying convention, he lumps all hominins in the past 1.8 million years into a single species, *Homo sapiens*. This will be too large a lump for most palaeoanthropologists (and me) to swallow. But his focus on hominin lineages — rather than individual species — has some merit, given that genetic evidence has arisen in recent years of ancient admixture between early modern humans, Neanderthals and Denisovans, and of gene flow into Denisovans from an unknown archaic hominin.

An admirable feature of Finlayson's hypothesis is its amenability to scientific testing. Further empirical data from fossils, artefacts and environmental records will certainly be valuable, as might a more nuanced treatment of individual sites. I would have welcomed more information about the chronology of these sites (which number more than 400), and the processes of burial, weathering and preservation that affected them over time. Sites close to water and in caves can be overrepresented in surveys, because their preservation potential is often higher than that of desert sites.

Ecological models are another avenue worth exploring. Mathematical models that examine foraging patterns and responses to habitat fragmentation in space and time, and simulations of alternative hypothetical scenarios, can help to illuminate what may have happened in the past. Such models could indicate which combinations of factors were likely to have had the greatest effect on hominin evolution. Multiple selective pressures have been in play over the past 7 million years and across the six inhabited continents, so untangling these interactions will be no easy task. *The Improbable Primate* provides a useful starting point for this next great challenge. ■

Richard G. Roberts *is Australian Research Council Laureate Fellow and director of the Centre for Archaeological Science at the University of Wollongong, Australia.*
e-mail: rgrob@uow.edu.au



Ocean currents swirl and eddy in a visualization of NASA satellite data, on show at the British Library.

INFOGRAPHICS

Truth is beauty

Daniel Cressey views the British Library's first science exhibition — a celebration of scientific illustration.

Isaac Newton may have long presided over the British Library in the form of Eduardo Palaozzi's vast sculpture at the library's entrance, but only now is the London institution hosting its first science exhibition. *Beautiful Science* catalogues attempts to make sense of the world through visualizations from the seventeenth century to today, drawing on the vast archives of the United Kingdom's national library. On show are graphics from alchemist Robert Fludd's 1617 work *Great Chain of Being* — which attempts to explain the Universe from stars to animals, vegetables and minerals — to a huge collection of modern Circos diagrams used to visualize genetic information and highlight relationships between species.

"Infographics are now a staple of every newspaper in the country. In many ways this seems like a new phenomenon," says curator Johanna Kieniewicz. "What I was really keen to show is that it actually has a very interesting and rich history."

The graphics have a many-layered power. "The visual representation of science can increase both the engagement of fellow researchers [and] the public," says Kieniewicz.

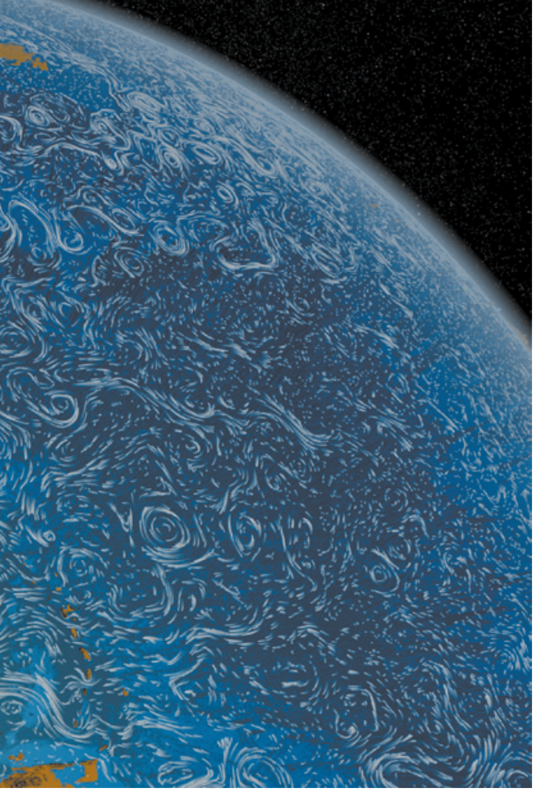
➲ NATURE.COM
For a video about the exhibition, see:
go.nature.com/cvy8r9

Beautiful Science: Picturing Data, Inspiring Insight
British Library, London. Until 26 May 2014.

Perhaps the exhibition's most famous expression of this is *Nightingale's Rose*, from the 1850s. A pioneer of modern medicine, Florence Nightingale demonstrated the value of improved hospital hygiene by showing that during the Crimean War, more British soldiers died as a result of poor sanitation in hospital than from enemy action. The image she drew to illustrate this point, says Kieniewicz, has "changed science, changed the way in which things are done".

Less transformative but no less impressive is William Farr's failed 1848–49 attempt to determine the cause of a cholera epidemic. His huge *Temperature and Mortality of London* plots these two variables in circular graphs. Farr's contemporary, John Snow, was more successful in using epidemiological mapping to pin cholera down as a waterborne disease; yet Farr's diagram stands as a monument to the difficulties of trying to tease causation out of huge data sets. Farr eventually came around to Snow's views, in part thanks to Snow's data presentation.

Beautiful Science shows that good data presentation is timeless. Witness Luke Howard, the meteorologist who named the



NASA/SVS

three basic cloud types — cirrus, cumulus and stratus. His diagram of the weather recorded in 1815 at his home in Tottenham, north London, sits next to a strikingly similar display of weather data from 2011. The latter is given a modern twist through an overlay of social-media commentary by design firm Clever°Franke of Utrecht, the Netherlands.

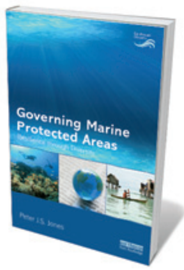
As science generated more complicated data, the methods needed to visualize them also became more complicated. Evolutionary theorist Jean-Baptiste Lamarck's 1809 table of the relationships between mammals, fish, reptiles, crustaceans and other animals can be understood easily by anyone with the most basic knowledge of French; much more parsing is needed for an avian tree of life published in *Nature* in 2012 that sits next to it. The exhibition reveals how seriously scientists need to take data presentation to convey meaning. Most of the pieces displayed were arduously plotted and coloured by hand. Says Kieniewicz of today's spreadsheet tools for chart production: "They make it too easy. You don't end up putting much thought into graphs."

Modern scientists, although liberated from painstaking drafting and shading, can gain from thinking hard about their audience, and about precisely what and how they want to communicate. A graph in a research paper today could be at the British Library in years hence, forming part of what Kieniewicz calls scientists' "visual legacy".

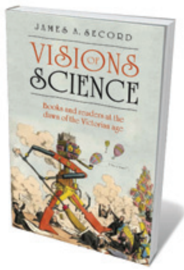
Science, the exhibition reminds us, is all about beautiful ideas. The trick is finding the aesthetic that maintains both meaning and elegance. ■

Daniel Cressey *is a reporter for* Nature *in London.*

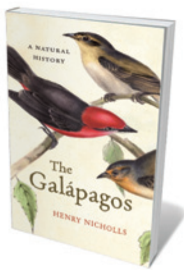
Books in brief



Governing Marine Protected Areas: Resilience through Diversity
Peter J. S. Jones Routledge *(2014)*
Marine Protected Areas, or MPAs — ocean zones limiting human activity — cover little more than 2% of the world's oceans, despite an internationally agreed target of 10% by 2020. And thousands of those that do exist are little more than 'paper parks', many scientists have found. Entering these choppy waters is geographer Peter Jones, who shows, through some 20 case studies, how and how not to govern MPAs effectively. Jones compellingly concludes that a diversity of incentives, from economic to social, is as essential as the diversity of the ecosystems MPAs are designed to protect.



Visions of Science: Books and Readers at the Dawn of the Victorian Age
James Secord Oxford University Press *(2014)*
Angst over scientific literacy is nothing new, notes James Secord. The political unease and religious turmoil in early Victorian Britain prompted John Herschel and fellow scientific utopians to urge a corrective: the dissemination of 'useful', or scientific, knowledge across society. As Secord shows, access to texts such as Charles Babbage's searing 1830 *Reflections on the Decline of Science in England* helped to foment an intellectual revolution in step with those in industry and modern science.



The Galapagos
Henry Nicholls Profile Books *(2014)*
From the plangent cries of blue-footed boobies to the plash of swimming iguanas, the Galapagos islands remain pristine — but for how long? In this natural and human history of Darwin's living laboratory, Henry Nicholls surfs from geology, oceanography and marine biology to resident land species, not least the burgeoning population of *Homo sapiens*. Throughout, he intertwines key accounts such as Darwin's inspired musings on geological uplift and the piscine encounters of pioneer diver William Beebe. One for the scientific islomane with a sense of the bigger picture.

The Rise of Superman: Decoding the Science of Ultimate Human Performance
Steven Kotler New Harvest *(2014)*
In extreme athletics, 'redefining the possible' is happening at an unprecedented rate. In this high-octane study, Steven Kotler explores 'flow', a neurochemically rich state in which cognitive and physiological processes mesh. The stupendous physical feats of the late ski-base jumper Shane McConkey and others are riveting. Equally surprising is what we know of flow science, such as how the brain's superior frontal gyrus deactivates to speed decision-making — and how 'bliss addiction' can result from too much exposure to flow states.

A Garden of Marvels: How We Discovered That Flowers Have Sex, Leaves Eat Air, and Other Secrets of Plants
Ruth Kassinger William Morrow *(2014)*
After 'murdering' a kumquat tree through hard pruning, Ruth Kassinger set out to understand plant physiology through the history of botany. In this scientific wonderland, we encounter Robert Hooke's exquisite drawings of plant structure in *Micrographia* (1665); arsenic-gobbling brake ferns; and the single eukaryote that, 1.6 billion years ago, engulfed a cyanobacterium and spawned the multitude of chloroplasts that made life on Earth possible. **Barbara Kiser**

Correspondence

Adapt current tools for handling big data

To speed up discoveries of disease biomarkers and treatments, we must work out a cheaper and faster way to process, store and use the huge medical data sets that are rapidly becoming available (see, for example, *Nature* **506,** 144–145; 2014).

By 2015, it is likely that a typical hospital will create 665 terabytes of data a year (for comparison, the web archive of the US Library of Congress contains less than 500 terabytes). This information can be used to study and analyse treatments — for example, for tuberculosis and stroke — and to reduce health-care costs.

To handle such big data effectively (see also *Nature* **498,** 255–260; 2013), we need to adapt classical information-processing tools. One computational challenge is how to manage the huge volume of detailed material as it becomes available, without sacrificing information. Another is that the data mostly represent physiological processes, the characteristics of which change over time.

Current methods are also inadequate for analysing collective information from different sensors, such as multi-dimensional descriptions from electroencephalography or magnetic resonance imaging of interactions between brain regions.

Ervin Sejdić *University of Pittsburgh, Pennsylvania, USA.*
esejdic@ieee.org

Portugal's research funding is secure

As president of the Portuguese Foundation for Science and Technology (FCT), I disagree with André Levy's overly pessimistic view of prospects for young scientists in the country (*Nature* **506,** 159; 2014).

The cuts in PhD studentships and postdoctoral fellowships that Levy mentions have been partially offset by FCT-funded PhD programmes and a funding boost for the latest round of fellowships. Plenty of opportunities will arise for recruitment and retention of young scientists through the steady funding of research projects and centres by the FCT, and from increased international funding through programmes such as the European Union's Horizon 2020 scheme.

In the past two years, the FCT has launched initiatives to strengthen Portugal's research base: PhD programmes and career-development schemes; differentiated research grants; a rigorous evaluation of centres of excellence; and top-up incentives for centres that bring in European or non-public funding. We are also building a national road map for research infrastructure.

The exponential growth of research and development in Portugal in the past 20–30 years is set to continue as the challenges of quality and international competitiveness are addressed.

Miguel Seabra *FCT, Lisbon, Portugal.*
presidencia@fct.pt

Forests are more than sticks of carbon

Valentin Bellassen and Sebastiaan Luyssaert suggest that forest managers should improve both carbon stocks and timber harvests to mitigate climate change (*Nature* **506,** 153–155; 2014). But forests are more than just sticks of carbon.

Maximizing harvests may be appropriate in heavily managed plantations. But in the remaining 96% of the world's forests, it could conflict with other forest uses and ecosystem services, as well as with biodiversity.

Some 1.6 billion people depend on forests to live. So optimizing wood production and carbon stocks at the expense of traditional uses and access rights could backfire. These people should not become carbon refugees.

The authors recommend "protecting trees from animals". In the African and Asian tropics particularly, such a move would have to consider the needs of elephants, rhinoceroses and ungulates, for example, to avoid undermining conservation efforts and the crucial ecosystem functions that these animals provide, such as seed and nutrient dispersal.

There is also a risk that simplifying forests to optimize two ecosystem services would reduce forest resilience to pest outbreaks and droughts, and to climatic and environmental changes (R. F. Noss *Conserv. Biol.* **15,** 578–590; 2001).

Promoting areas of multi-use forest that have various functions would provide environmental and social benefits, while retaining more carbon than smaller, single-use areas.

Aerin L. Jacob, Sarah Jane Wilson *McGill University, Montreal, Canada.*
aerin.jacob@mail.mcgill.ca
Simon L. Lewis *University College London, UK.*

Focus on questions, not hypotheses

I contend that the insistence of the US National Institutes of Health (NIH) on hypothesis-driven projects in grant proposals could be a factor contributing to irreproducible research reports (see F. S. Collins and L. A. Tabak *Nature* **505,** 612–613; 2014).

Isaac Newton argued that "hypotheses … have no place in experimental philosophy", a view echoed by mathematician Roger Cotes: "Those who assume hypotheses as first principles of their speculations … may indeed form an ingenious romance, but a romance it will still be" (in I. B. Cohen *Introduction to Newton's* Principia; iUniverse, 1999).

Such criticisms recognize the risk that scientists may filter data through their hypotheses, discounting results that do not validate the hypothesis as evidence that the experiment did not work — rather than as evidence that the hypothesis is false.

The NIH's funding criteria should instead ensure that a pertinent research question is being asked, and that the applicant has the means to answer it (see D. J. Glass *Experimental Design for Biologists*; Cold Spring Harbor Laboratory Press, 2006).

David J. Glass *Novartis Institutes for Biomedical Research, Cambridge, Massachusetts, USA.*
david.glass@novartis.com

Update forensics for deaths in Japan

Autopsies in Japan are performed on only 1.6% of all deaths, the lowest rate among developed countries. Cause of death is almost always determined by simple visual inspection, as it was 100 years ago. Japanese forensic medicine needs to catch up to avoid murders being overlooked.

In 2011, the National Police Agency announced that 43 murders had been missed since 1998, with many found to have been disguised as suicides or natural deaths.

The agency is now trying to carry out more autopsies, but examinations are often cursory because of budget restrictions. To tackle the problem, rigorous autopsies need to be performed routinely.

In 2012, two laws relating to death investigations were passed, and the government established the Committee for the Promotion of Cause of Death Investigation, which published interim reports in May 2013. No governmental agency will take responsibility for the budget, however, so little progress has been made.

Exposing the possible involvement of crime in unnatural deaths continues to be as hard as it ever was in Japan.

Hiroshi Ikegaya *Kyoto Prefectural University of Medicine, Kyoto, Japan.*
ikegaya@koto.kpu-m.ac.jp

SPACE SCIENCE

Near-Earth space shows its stripes

Electrons in Earth's magnetic field often exhibit a striped pattern of intensity as a function of electron energy and altitude. A model that factors in some unexpectedly important effects can now explain this feature. SEE LETTER P.338

DREW L. TURNER

Using some of the first scientific satellites put into orbit during the late 1950s, teams led by physicists James Van Allen in the United States and Sergei Vernov in the Soviet Union independently reported[1,2] on defined regions of radiation in near-Earth space. These regions came to be known as Earth's radiation belts, and they represent the first major scientific discovery of the space age. However, despite decades of study, many questions in radiation-belt physics remain unanswered, mostly concerning the nature of the inner and outer belts, which are populated by electrons moving at near the speed of light. As society becomes ever more dependent on satellite-based technology, it is increasingly important to understand the variability in the radiation belts, because the highest-energy "killer electrons"[3] can result in potentially fatal damage to sensitive spacecraft electronics[4]. On page 338 of this issue, Ukhorskiy *et al.*[5] present observations and a model of a previously unexplained structured feature of the inner radiation belt, which they call zebra stripes.

Examining early mission data from instruments for RBSPICE (the Radiation Belt Storm Probes Ion Composition Experiment)[6] on board NASA's twin Van Allen Probes satellites[7], Ukhorskiy and colleagues found the distinct zebra-stripe pattern in plots of electron intensity as a function of electron energy and equatorial radial distance, an example of which is shown in Figure 1a. Such a pattern had been inferred[8] in observations from the low-altitude Demeter spacecraft and was thought to be caused by interactions between inner-belt electrons and ultralow-frequency plasma waves during geomagnetic storms — active periods of geomagnetic activity.

However, with the new observations from RBSPICE, which observed the stripes on most orbital passes by the satellites, it became clear that the stripes are a regular feature of the inner radiation belt. They were observed within an altitude of about 13,000 kilometres at low geomagnetic latitudes (latitudes with respect to Earth's magnetic axis), even during quiet geomagnetic conditions. Contradictory to the previous hypothesis, the stripes are actually more coherent and stable during periods of low geomagnetic activity, inspiring Ukhorskiy *et al.* to re-evaluate how these structured features might form.

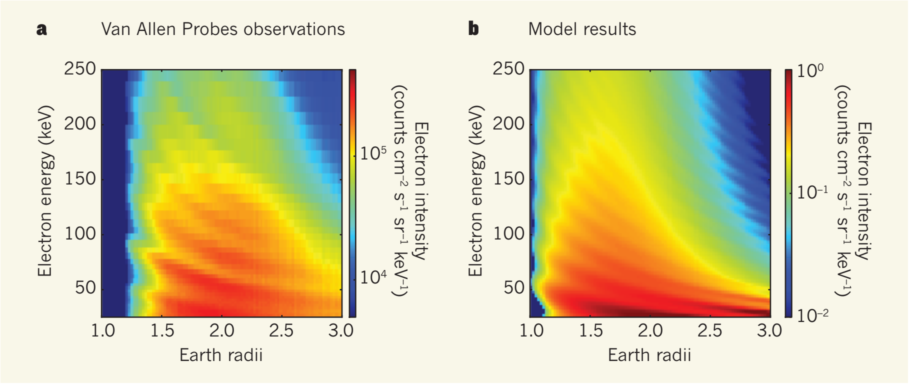


Figure 1 | Zebra-stripe patterns. a, An example of observations of zebra-stripe patterns in the intensity of electrons in Earth's inner radiation belt as a function of electron energy (in kiloelectronvolts) and distance from Earth (expressed in Earth radii from the centre of the planet). The observations were made by the RBSPICE instruments on board the twin Van Allen Probes satellites. Intensity is shown in units of electron number (counts) per square centimetre per second per steradian per keV. **b**, Ukhorskiy *et al.*[5] reproduce these observations using a model in which drifting electrons in an Earth-like magnetic field interact resonantly with an electric field induced by a tilt between the planet's magnetic moment and its axis of rotation. (Plots modified from Fig. 1 of the paper[5].)

The dynamics of energetic charged particles in Earth's magnetic field, such as electrons and protons, are dominated by a combination of electric and magnetic forces called the Lorentz force, and, because of the mainly dipolar nature of the field, radiation-belt electrons can become trapped in much the same manner as in the classical field arrangement known as a magnetic 'bottle'. The trapped particles undergo three types of periodic motion: gyration around magnetic-field lines; bounce motion along field lines, potentially from pole to pole; and drift motion perpendicular to field lines in the azimuth (longitudinally) around the planet, owing to forces associated with the gradient and curvature of the magnetic field.

Furthermore, because Earth's magnetic moment is not parallel to its axis of rotation, an induced electric field is also generated. This oscillating electric field was thought to be unimportant for radiation-belt dynamics at Earth, because the kinetic energy of radiation-belt particles is many orders of magnitude stronger than that imposed by the electric field. However, Ukhorskiy and colleagues speculate that, because Earth's rotation is independent of geomagnetic activity, the electric field might have a role in the formation of the zebra stripes.

The researchers used a model that simulates the dynamics of test particles in an Earth-like magnetic field and that includes the induced, oscillating electric field. The model shows that initially smooth distributions of electrons evolve into zebra-stripe patterns (Fig. 1b). Such patterns form because of an interaction (phase resonance) between the drift motions of electrons, which are related to their energies and radial locations, and the diurnal oscillations of the induced electric field. In this interaction, the electrons undergo a phenomenon called drift phase mixing, which affects their energy and radial location, resulting in distinct peaks and troughs in electron intensity as a function of energy and radial location. Thus, the zebra stripes in the observed electron intensities are simply manifestations of the electron distributions produced by such drift phase mixing.

Ukhorskiy and colleagues' model is relatively simple, yet it simulates the zebra stripes remarkably well during quiet geomagnetic conditions. The authors also demonstrate that the model can be modified to explain how the zebra-stripe patterns are disrupted by perturbations during geomagnetic storms.

This new understanding will compel space physicists to re-evaluate their concept of radiation-belt dynamics and its finer complexities. The results also shed light on the nature of radiation-belt dynamics in other planetary systems[9], such as those of Jupiter, Neptune and Uranus, because these planets all have a fast rotation period and significant tilt between the magnetic moment and axis of rotation (unlike Saturn). We can look forward to testing the model using observations for these systems from future missions. ■

Drew L. Turner *is in the Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, Los Angeles, California 90095-1567, USA.*
e-mail: dturner@igpp.ucla.edu

1. Van Allen, J. A. & Frank, L. A. *Nature* **183,** 430–434 (1959).
2. Vernov, S. N. & Chudukov, A. E. in *Proc. Moscow Cosmic Ray Conf.* Vol. 3 (ed. Syrovatsky, S. I.) 19–29 (Int. Un. Pure Appl. Phys., Moscow, 1960).
3. Horne, R. B. *Nature Phys.* **3,** 590–591 (2007).
4. Baker, D. N. *Science* **297,** 1486–1487 (2002).
5. Ukhorskiy, A. Y. *et al. Nature* **507,** 338–340 (2014).
6. Mitchell, D. G. *et al. Space Sci. Rev.* **179,** 263–308 (2013).
7. Mauk, B. H. *et al. Space Sci. Rev.* **179,** 3–27 (2013).
8. Sauvaud, J.-A. *et al. J. Geophys. Res. Space Phys.* **118,** 1723–1736 (2013).
9. Mauk, B. H. & Fox, N. J. *J. Geophys. Res.* **115,** A12220 (2010).

GENETICS

Closing the distance on obesity culprits

Genetic variation in a non-protein-coding region of the gene *FTO* is implicated in obesity. A study finds evidence that, rather than affecting *FTO* itself, variations in this region influence expression of a distant gene, *IRX3*. **See Letter p.371**

DAVID U. GORKIN & BING REN

Genome-wide association studies have identified more than 75 locations in the human genome at which changes in DNA sequence are linked to obesity and traits closely related to the condition, such as body mass index. In these studies, single nucleotide changes in a non-protein-coding region (an intron) of the gene *FTO* often have the strongest association with obesity[1–3]. It is generally thought that this association arises because of a genetic defect that alters the function or expression level of *FTO*, which encodes an enzyme known to be involved in the control of body weight and metabolism in mice[4,5]. However, evidence of a direct link between the single nucleotide changes and alteration of *FTO* function or expression level has been lacking. In this issue, Smemo *et al.*[6] (page 371) present compelling evidence that the association between obesity and *FTO* actually involves another gene called *IRX3*, which encodes a transcription factor involved in multiple developmental processes.

IRX3 is located a long way — roughly 500 kilobases — from *FTO*'s first intron, in which the single nucleotide changes (also known as single nucleotide polymorphisms, or SNPs) associated with obesity are located. Despite previous research[7] suggesting that long-range enhancer sequences in *FTO*'s first intron might regulate the expression of *IRX3*, attention has continued to focus on *FTO*, partly owing to a lack of more-definitive evidence linking these enhancers to *IRX3* expression. Now, Smemo and colleagues have confirmed and expanded on this link.

Verifying a functional relationship between an enhancer sequence and a distant target gene requires a demonstration that the enhancer interacts physically with the target gene's promoter in the cell nucleus (a promoter is a DNA sequence immediately adjacent to the beginning of a gene that is involved in regulating transcription). Smemo and colleagues investigated long-range interactions between *FTO* and *IRX3* using a technique called chromatin conformation capture, which measures how frequently different regions of the genome interact[8]. The authors found that the promoter of *IRX3* physically interacts with the first intron of *FTO* (Fig. 1). This interaction was found in mouse and zebrafish embryos, adult mouse brains and human cell lines, suggesting that interaction between these two regulatory regions is an evolutionarily conserved feature of genome organization. The authors also found that sequences in the first intron of *FTO* show enhancer activity in several of the tissues in which *IRX3* is expressed, and that the promoter of *IRX3* alone cannot account for the gene's observed expression pattern, indicating that *IRX3* relies on long-range regulatory input.

Smemo and co-workers reasoned that, if the association between obesity and *FTO* does result from genetic alteration of an *IRX3* enhancer, then the same SNPs linked with obesity should also be associated with changes in *IRX3* expression. To investigate this, the authors examined *IRX3* expression in 153 human brain samples. Obesity-linked SNPs were associated with *IRX3* expression in these samples, but not with expression of *FTO*, directly linking these variants to *IRX3* regulation.

In a final, but crucial, set of experiments, Smemo *et al.* found that mice lacking a functional copy of *Irx3* have significantly lower body weight than control animals, along with a higher base metabolic rate, less body fat and a predisposition to form brown over white adipose tissue (the latter being more likely to accumulate in obese individuals). This clearly demonstrates a previously unappreciated role for *Irx3* in controlling body weight. Furthermore, the authors report that these same traits are also seen in mice when a 'dominant-negative' mutant form of *Irx3*, which interferes with the function of normal *Irx3*, is expressed in the hypothalamus, a region of the brain that has a key role in regulating food intake and metabolism. These data indicate that control of body weight relies specifically on *Irx3* activity in the hypothalamus.

Smemo and colleagues' data suggest a model in which a particular genetic variant (or variants) in the first intron of *FTO* disrupts an enhancer of *IRX3,* thus altering *IRX3* expression and, in turn, influencing body weight. However, important questions remain. First, what is the true functional variant (or variants) that causes the association between this region

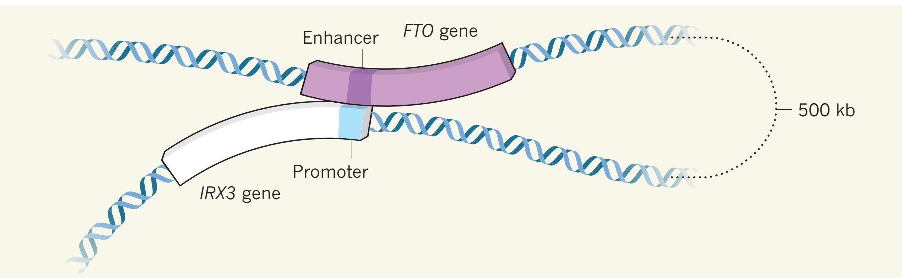


Figure 1 | Long-range control of gene expression. Smemo *et al.*[6] have demonstrated that a non-protein-coding region of the gene *FTO*, which has been linked to obesity, physically interacts with the promoter sequence required for expression of the *IRX3* gene, located around 500 kilobases (kb) away. The authors find that sequences within this non-protein-coding region of *FTO* act as enhancer elements, regulating expression of the distant *IRX3*.

and obesity? Although many SNPs identified in this region have been associated with obesity, it is likely that these associations reflect a single genetic defect in the region. Without pinpointing this defect, it will be difficult to determine whether the genetic mechanism involves disrupted regulation of *FTO*, *IRX3* or perhaps both. Second, what tissues are most relevant to this association? The current study does an admirable job of examining appropriate tissues wherever possible, but there are missing links. For example, the authors observed that loss of *Irx3* function in the hypothalamus causes changes in body weight in mice, but the human brain samples were from the cerebellum. Additional studies are needed to answer these and other questions about the connection between obesity, *FTO* and *IRX3*.

The implications of Smemo and co-workers' study extend far beyond obesity. SNPs associated with other human traits frequently occur at or near transcriptional enhancers[9], and linear genomic distance is often an unreliable predictor of enhancer–promoter interactions[10,11]. The connection between *IRX3* and *FTO* is particularly instructive because the enhancers of one gene reside within an intron of the other. However, this is not an isolated case. For example, an enhancer in an intron of the gene *LMBR1* regulates the developmental gene *SHH* located more than 1 megabase away, and mutations in this enhancer can cause limb malformations owing to altered *SHH* expression[12]. Similarly, an enhancer within an intron of the gene *HERC2* contains a SNP that influences human eye, hair and skin colour by modulating the expression of a neighbouring gene called *OCA2* (ref. 13). Through their study, Smemo *et al.* provide an excellent example of how detailed functional analyses, particularly those that link enhancers to their target genes, can help to evaluate the biological consequences of variation in non-coding sequences. ■

David U. Gorkin *and* **Bing Ren** *are at the Ludwig Institute for Cancer Research, La Jolla, California 92093, USA.* **B.R.** *is also in the Department of Cellular and Molecular Medicine, University of California, La Jolla, USA.*
e-mails: dgorkin@gmail.com; biren@ucsd.edu

1. Frayling, T. M. *et al. Science* **316,** 889–894 (2007).
2. Scuteri, A. *et al. PLoS Genet.* **3,** e115 (2007).
3. Dina, C. *et al. Nature Genet.* **39,** 724–726 (2007).
4. Fischer, J. *et al. Nature* **458,** 894–898 (2009).
5. Church, C. *et al. Nature Genet.* **42,** 1086–1092 (2010).
6. Smemo, S. *et al. Nature* **507,** 371–375 (2014).
7. Ragvin, A. *et al. Proc. Natl Acad. Sci. USA* **107,** 775–780 (2010).
8. van de Werken, H. J. G. *et al. Nature Methods* **9,** 969–972 (2012).
9. Maurano, M. T. *et al. Science* **337,** 1190–1195 (2012).
10. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. *Nature* **489,** 109–113 (2012).
11. Jin, F. *et al. Nature* **503,** 290–294 (2013).
12. Lettice, L. A. *et al. Hum. Mol. Genet.* **12,** 1725–1735 (2003).
13. Visser, M., Kayser, M. & Palstra, R.-J. *Genome Res.* **22,** 446–455 (2012).

This article was published online on 12 March 2014.

TECHNOLOGY

Photonics illuminates the future of radar

The first implementation of a fully photonics-based coherent radar system shows how photonic methods for radio-frequency signal generation and measurement may facilitate the development of software-defined radar systems. SEE LETTER P.341

JASON D. MCKINNEY

For well over a decade, there has been intense pursuit of photonic techniques for the synthesis and processing of broadband radio-frequency signals in applications ranging from wireless communications to radar. Such work is motivated by one main challenge: the lack of electronic systems capable of directly generating, processing and digitizing signals that have high central frequencies and large bandwidths while maintaining signal fidelity (minimal spurious-signal content and low noise levels). That challenge is sure to remain, given the escalating interest in higher centre frequencies and greater signal agility for software-defined radio and radar architectures — those in which the desired signal can be rapidly changed under software control. Such signal agility could offer, for example, increased radar resolution. On page 341 of this issue, Ghelfi *et al.*[1] describe the first field demonstration of a coherent radar system based on photonic techniques. Their work shows that such techniques may indeed provide capability for next-generation radar systems.

Coherent radar systems[2] use both the amplitude and phase (frequency) of the radar signal that is reflected from a given target (the radar return) to provide range and velocity data for the target. In such a system, a radar waveform of low centre frequency is designed — that is, its bandwidth is chosen — to give the desired range resolution. This waveform is then shifted to a higher centre frequency by multiplication with a stable, continuous-wave radio-frequency (RF) signal (carrier), and subsequently amplified and transmitted to the target. On reaching the radar receiver, the waveform is shifted to a lower, intermediate frequency, again through multiplication with a stable RF signal, and processed to extract the target's range and velocity.

Information on range is acquired from the time of arrival of the radar return. The range resolution is determined by the waveform's bandwidth-limited duration, and improves as this duration decreases. The need for increased resolution requires the use of higher carrier frequencies, which are necessary to obtain waveforms of shorter duration. Velocity data are obtained by comparing the frequency of the return waveform against a stable, continuous-wave reference signal of the same frequency as that of the original radar waveform. For a moving target, the return waveform will be shifted in frequency relative to its original frequency owing to the Doppler effect (at acoustic frequencies, this effect is analogous to the way in which the perceived pitch of a

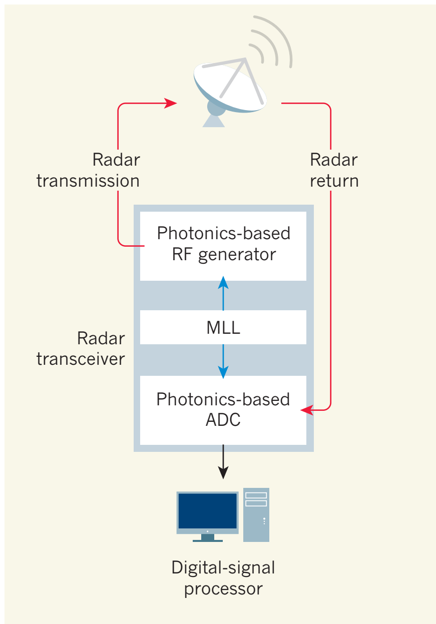


Figure 1 | A photonics-based radar system. At the heart of the radar transmitter–receiver (transceiver) implemented by Ghelfi *et al.*[1] is a mode-locked laser (MLL). This laser is connected by optical fibres (blue arrows) to a photonics-based radio-frequency (RF) generator, in which the laser output is first optically modulated and then converted to electronic form to obtain a radar signal. The MLL is also connected to a photonics-based analog-to-digital converter (ADC). The generator's output is connected to an antenna by an RF cable (left red arrow) and transmitted towards a target. The signal reflected from the target (radar return) is collected by the antenna and directed by means of the RF cable (right red arrow) to the ADC. The ADC uses a train of pulses from the MLL to optically measure the amplitude of the reflected signal at discrete instances in time with high precision (optical sampling). The resulting optical samples are converted to digital electronic form and transmitted (black arrow) to the digital-signal processor of a computer. (Graphic adapted from Fig. 2 of the paper[1].)

siren varies as it passes an observer). Accurate velocity measurement demands that the frequency of both the carrier and the reference signals be known with a high degree of certainty — that is, that their phase noise is low.

In a digital coherent radar, processing of the radar return is accomplished using a computer. Such processing requires the use of a high-precision analog-to-digital converter (ADC), which measures the amplitude of the return waveform at discrete instances in time and converts these measurements (samples) to digital data suitable for computer processing. This digitization requires that the time at which each sample is taken in the ADC is known to high accuracy, which in turn necessitates that the timing jitter — the variation in the time between successive samples — is minimal. Low phase noise and minimal timing jitter are intimately related: timing jitter is directly proportional to the phase noise mathematically integrated in frequency[3]. Minimizing the phase noise therefore minimizes the timing jitter.

Enter the mode-locked laser — a class of laser in which the optical modes of oscillation maintain a highly stable phase relationship (hence the term mode-locked). This phase stability results in a periodic train of laser pulses exhibiting low timing jitter. When two optical modes are selected by an optical filter placed beyond the laser and subsequently directed to a light sensor known as a photodiode, the sensor's electrical output is an RF signal with low phase noise. The frequency of this signal is readily tuned on the basis of the two optical modes that are selected, and the efficiency with which the RF signal is created is determined by the electrical response of the photodiode at the chosen frequency. For coherent-radar applications, such a laser provides an appealing tool that can both generate tunable low-phase-noise RF signals and serve as the basis for a low-jitter ADC. This is precisely what Ghelfi *et al.* have done to build a coherent radar system as part of a project called PHODIR ('Photonics-based fully digital radar').

In this project, a single mode-locked laser is used to generate the required continuous-wave RF signals and to sample the return waveforms in an optical ADC. Such operations are at the centre of the project's photonics-based radar transmitter–receiver, or transceiver (Fig. 1). Laboratory measurements of the resulting continuous-wave signals showed that signals with a frequency near 40 gigahertz exhibit integrated phase noise and timing jitter that are approximately half those of state-of-the-art conventional RF synthesizers. More impressively, because the timing jitter of the laser-pulse train is about ten times lower than that of the best electronic ADCs, the authors achieved record-precision digitization of continuous-wave signals at 40 GHz. This optical ADC has about 100 times better digitization fidelity than its best electronic equivalent. And the device showed comparable digitization fidelity to that of similar photonic ADCs[4] over a bandwidth that was ten times larger — a remarkable result.

In a first-of-its-kind demonstration, Ghelfi and colleagues next incorporated their photonics-based transceiver into a coherent radar architecture operating at a carrier frequency of about 10 GHz, and tested it in a real-world situation by measuring the take-off trajectories of several aircraft. They compared the resulting range and velocity data against publicly available trajectory data from independent sources and found excellent agreement between the two data sets. The results illustrate that a photonics-based transceiver can be successfully used in a radar system.

The individual performances of the system's transmitter and ADC are world-class with respect to those of other photonics-based devices. Furthermore, the readily tunable carrier and reference signals, as well as the photonic ADC, remove much of the complexity of frequency translation that is inherent in all-electronic radar implementations. Certainly, these elements, and their combination into a functional system, are appealing components for future frequency-agile, software-defined radar architectures.

But an important question must be addressed: can photonics-based radars achieve similar or better performance than their all-electronic counterparts? Answering this question will require direct comparison of photonic and all-electronic versions of the same radar. Compared with the system described here, current commercial digital radar systems operating at a 10-GHz carrier frequency show substantially higher dynamic range (the range of received-signal power over which the radar receiver responds linearly). Because dynamic range translates directly into radar sensitivity and false-target detection, a high value for this quantity is crucial.

Another concern is one of radar range, the distance over which the radar can successfully detect a target. Increased range requires longer processing time, placing more stringent limits on long-term timing jitter and phase noise at a low frequency offset from the carrier signal — quantities that are difficult to minimize[5] in photonics-based architectures.

Ghelfi and colleagues' study raises yet other questions. In what types of radar system could photonics have the most impact? Given a particular radar application, what requirements (associated with, for example, timing jitter and phase noise) are placed on our techniques, and can they be achieved? Pursuing the answers to these questions will no doubt lead to exciting — and challenging — research in the years to come. ■

Jason D. McKinney *is at the Microwave Photonics Section, Optical Sciences Division, US Naval Research Laboratory, Washington DC 20375, USA.*

e-mail: jason.mckinney@nrl.navy.mil

1. Ghelfi, P. *et al. Nature* **507,** 341–345 (2014).
2. Skolnik, M. (ed.) *Radar Handbook* 2nd edn (McGraw-Hill, 1990).
3. von der Linde, D. *Appl. Phys. B* **39,** 201–217 (1986).
4. Kim, J., Park, M. J., Perrott, M. H. & Kärtner, F. X. *Opt. Express* **16,** 16509–16515 (2008).
5. Fortier, T. M. *et al. Nature Photon.* **5,** 425–429 (2011).

CELL BIOLOGY

The disassembly of death

During the cell-death program known as apoptosis, cells break up into membrane-bound fragments. It emerges that this process is controlled by the protein pannexin 1 and can be deregulated by an antibiotic. **SEE ARTICLE P.329**

CHRISTOPHER D. GREGORY

Cell death is a feature of normal life. Around the clock, cells are disassembled by an evolutionarily conserved controlled-death program called apoptosis, which ensures proper regulation of the size and quality of cell populations in tissues. Much is known about the molecular mechanisms that trigger this program, but it is unclear how dying cells are dismantled in a controlled manner once apoptosis is under way. On page 329 of this issue, Poon *et al.*[1] demonstrate a central role for a cell-membrane channel, the pannexin 1 protein, in inhibiting cell disassembly during apoptosis, and report that, unexpectedly, a member of a commonly prescribed class of antibiotic is able to modulate the activity of such channels.

Pannexin 1 (PANX1) channels are complex structures that span the plasma membrane of cells, regulating the movement of small molecules of various sizes between the intracellular and extracellular milieux. These channels are best known for controlling the release of ATP, a multifunctional energy-carrying and signalling molecule. It was previously shown[2,3] that PANX1 channels seemingly open up to release ATP from the cell during apoptosis. This release acts as a 'find-me' signal for phagocytes — cells that seek out, engulf and digest apoptotic cells and their fragments (Fig. 1a). In addition to ATP, PANX1 channels are likely to control the efflux and influx of a wide variety of bioactive molecules from many cell types in various states of activation and differentiation, as well as cell death[4]. Conveniently, PANX1-channel opening can be monitored using surrogate markers, because fluorescent DNA-binding dyes such as TO-PRO-3 are among the molecules that selectively enter cells through these channels[1,3].

In their groundbreaking paper, Poon and colleagues demonstrate that rigorous molecular screening, when coupled with the power of observation in a well-regulated setting, can uncover surprising and invaluable findings. The authors designed an assay in which uptake of TO-PRO-3 into apoptotic human cells was monitored as an indicator of PANX1-channel activity. They subjected this assay to a widely available library of 1,280 small molecules, including the quinolone-based antibiotic, trovafloxacin. Unexpectedly, trovafloxacin proved to be a potent and direct pharmacological inhibitor of PANX1 channels. The authors demonstrated that this role was independent of apoptosis, because trovafloxacin was able to inhibit individual, non-apoptotic channels.

Poon and co-workers went on to study the effects of trovafloxacin on apoptotic cells and discovered that PANX1 channels not only regulate ATP release during apoptosis, but also control the formation of apoptotic bodies, membrane-bound cell fragments that bud off from the cell during apoptosis. Inhibition of normal PANX1-channel function with trovafloxacin resulted in a large increase in the number of apoptotic bodies formed, causing dysregulated fragmentation of the cell (Fig. 1b). The authors also found that active PANX1 channels prevent the formation of string-like structures (which they term apoptopodia) that arise during inhibition of channel activity in apoptotic cells. Somehow apoptopodia prevent the detachment of blebs — irregular bulges in the plasma membrane — that commonly appear at cell surfaces during the early stages of apoptosis.

The purpose of apoptotic bodies is currently unknown. Each apoptotic body has a diverse mixture of components, and may contain DNA, fragments of the cell nucleus, other intact organelles and a wide array of intracellular proteins and other molecules, all encapsulated by portions of the membrane that previously surrounded the living cell. Although the bodies are rapidly 'gobbled up' by phagocytes, whole dying cells can also be engulfed effectively; in fact, not all apoptotic cells break up. Perhaps apoptotic bodies are

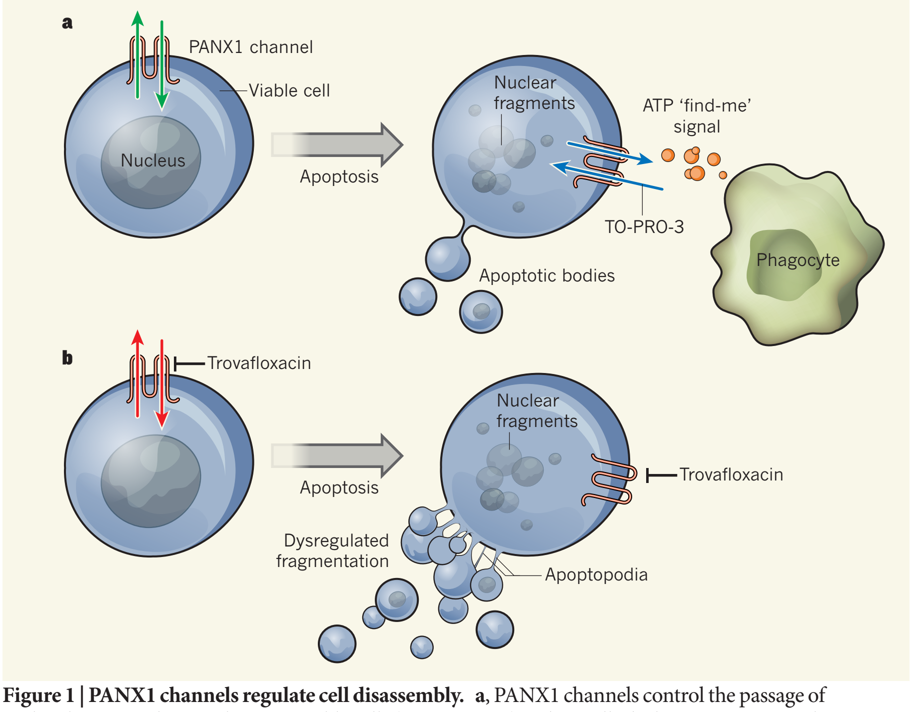


Figure 1 | PANX1 channels regulate cell disassembly. a, PANX1 channels control the passage of molecules across the membrane in viable cells (green arrows). When cells die by apoptosis, membrane-bound fragments called apoptotic bodies bud off from the cell. PANX1 channels release ATP molecules from the cell and allow entry of selected molecules such as the fluorescent dye TO-PRO-3 (blue arrows). ATP acts as a 'find-me' signal to phagocytes that engulf both the dying cell and the apoptotic bodies. **b**, Poon *et al.*[1] report that PANX1-channel function is inhibited by the antibiotic trovafloxacin, which modulates the channel's activity in viable cells (red arrows). Channel inhibition also results in the formation of many more apoptotic bodies during apoptosis, leading to dysregulated fragmentation of the cell. The study reveals that apoptotic bodies form in association with string-like structures, termed apoptopodia.

more efficiently degraded than whole apoptotic cells within the phagocyte, or maybe they are targeted or degraded differently. Apoptotic bodies may also be important in controlling immune responses to apoptotic cells[5].

It is possible that apoptotic bodies retain biological functions and act at sites well away from the dying cell. Given that apoptosis can drive cell regeneration in tissues[6] — an attractive replacement of the old with the new — it is particularly tempting to suggest that apoptotic-cell-derived bodies transport active components that contribute to these regenerative processes. Small vesicles have been found by others to export bioactive cargoes from non-apoptotic cells in response to various stimuli[7]. Although such vesicles are generally smaller than apoptotic bodies, the two types have certain structural characteristics in common. Exactly how PANX1 channels contribute to the functional (including possible pathological) attributes of apoptotic bodies, and perhaps other extracellular vesicles, will be valuable avenues of future investigation.

Poon and colleagues' work raises an issue of relevance to improved antibiotic design. Trovafloxacin, a highly effective, broad-spectrum fluoroquinolone antibiotic, was withdrawn because of adverse side effects in humans[8]. Fluoroquinolones in general represent some of the most commonly prescribed antibiotics, but they carry the risk of severe side effects that can harm multiple organ systems[9]. Could abrogating unwanted targeting of PANX1 channels by these drugs offer an opportunity to develop improved antibiotics? As yet, we do not know whether dysregulated production of apoptotic bodies has pathological consequences, or whether another, unidentified mechanism could cause adverse clinical effects when patients are treated with trovafloxacin — a mechanism that might be dependent or independent of PANX1-channel inhibition. What is crucial to determine is the functional diversity of PANX1 channels, and the pathological consequences of their blockade.

As always, the unexpected is the most fascinating. There is no doubt that this work will stimulate researchers in the seemingly disparate areas of apoptosis and antibiotics to find out more about PANX1. The results may improve not only scientific understanding but also health care. ■

Christopher D. Gregory *is at the MRC/University of Edinburgh Centre for Inflammation Research, Queen's Medical Research Institute, University of Edinburgh, Edinburgh EH16 4TJ, UK.*
e-mail: chris.gregory@ed.ac.uk

1. Poon, I. K. H. *et al. Nature* **507,** 329–334 (2014).
2. Elliott, M. R. *et al. Nature* **461,** 282–286 (2009).
3. Chekeni, F. B. *et al. Nature* **467,** 863–867 (2010).
4. Sandilos, J. K. & Bayliss, D. A. *J. Physiol. (Lond.)* **590,** 6257–6266 (2012).
5. Wickman, G., Julian, L. & Olson, M. F. *Cell Death Differ.* **19,** 735–742 (2012).
6. King, R. S. & Newmark, P. A. *J. Cell Biol.* **196,** 553–562 (2012).
7. Hugel, B., Martínez, M. C., Kunzelmann, C. & Freyssinet, J. M. *Physiology* **20,** 22–27, (2005).
8. Lucena, M. I. *et al. Clin. Infect. Dis.* **30,** 400–401 (2000).
9. Liu, H. H. *Drug Safety* **33,** 353–369 (2010).

This article was published online on 12 March 2014.

BONE BIOLOGY

Vessels of rejuvenation

The identification of specialized endothelial-cell populations in the blood vessels of bones, and their signalling pathways, reveals how the vasculature contributes to bone formation. SEE ARTICLE P.323 & LETTER P.376

FERDINAND LE NOBLE & JOS LE NOBLE

The different structural and metabolic requirements of organs and tissues prompt the formation of specialized blood vessels during development[1], which deliver molecules that instruct tissues on their fates[2,3]. This process requires crosstalk between organs and the vasculature: organs generate local growth factors that attract blood vessels toward metabolically active regions, and the vessels release molecules (known as angiocrine signals) that affect cell differentiation in the developing organs. It is becoming increasingly clear that these interactions are influenced by the endothelial cells that line the inner surface of blood vessels. In two papers[4,5] in this issue, Adams and colleagues demonstrate that the bone vasculature contains specialized endothelial cells with signalling properties that support bone maturation and regeneration. As well as providing insight into the tissue specificity of angiocrine signalling, the findings have direct clinical relevance.

In the first paper, Kusumbe *et al.*[4] (page 323) began by mapping gene expression, metabolic activity and cell-surface markers in the developing postnatal bone vasculature of mice. The authors distinguished two new subpopulations of endothelial cell (EC) that they term type H and type L ECs. Interestingly, they found that osteoprogenitor cells, which differentiate into

NATURE

50 Years Ago

The Agar Precipitation Technique and Its Application as a Diagnostic and Analytical Method by F. Peetoom — Nomenclature is not discussed critically, and the style is diffuse and difficult to read. A translation which makes no better attempt at English style than "After the eleven each lasting one minute incubations the serum hardly contained …", etc. (p. 88), the strange punctuation, and the various misprints should not have appeared from a British publishing house.
From *Nature* 21 March 1964

100 Years Ago

'An Elizabethan Cookery-book'
A Proper Newe Booke of Cokerye edited by Catherine F. Frere — The original Cambridge edition is dateless … the book was often reprinted before 1546, and was … the "Book of Cookery" of 1500 … As we turn over the leaves of the "Proper Newe Booke," with its quaint recipes, couched in the "corrupted phonetic" of the golden age of English prose, we gather…"a little rushlight illumination on the culinary mysteries of the once busy kitchens, roofless and empty to-day, and on the hospitalities, feasting, and revels of the now silent dining halls of long ago."

ALSO:

Prof. Arthur Keith's discourse on an anthropological study of some portraits of Shakespeare … is printed in … the *British Medical Journal* … Examination of the head of Shakespeare proves him to be a representative example of the short-headed type, not of the early British breed, but of the round-headed race which entered Britain in the Bronze period about 2000 B.C. His brain capacity was more than 1900 cm., as compared with 1477 in an average Englishman.
From *Nature* 19 March 1914

ARTICLE

OPEN

doi:10.1038/nature12965

Comprehensive molecular characterization of urothelial bladder carcinoma

The Cancer Genome Atlas Research Network*

Urothelial carcinoma of the bladder is a common malignancy that causes approximately 150,000 deaths per year worldwide. So far, no molecularly targeted agents have been approved for treatment of the disease. As part of The Cancer Genome Atlas project, we report here an integrated analysis of 131 urothelial carcinomas to provide a comprehensive landscape of molecular alterations. There were statistically significant recurrent mutations in 32 genes, including multiple genes involved in cell-cycle regulation, chromatin regulation, and kinase signalling pathways, as well as 9 genes not previously reported as significantly mutated in any cancer. RNA sequencing revealed four expression subtypes, two of which (papillary-like and basal/squamous-like) were also evident in microRNA sequencing and protein data. Whole-genome and RNA sequencing identified recurrent in-frame activating *FGFR3–TACC3* fusions and expression or integration of several viruses (including HPV16) that are associated with gene inactivation. Our analyses identified potential therapeutic targets in 69% of the tumours, including 42% with targets in the phosphatidylinositol-3-OH kinase/AKT/mTOR pathway and 45% with targets (including ERBB2) in the RTK/MAPK pathway. Chromatin regulatory genes were more frequently mutated in urothelial carcinoma than in any other common cancer studied so far, indicating the future possibility of targeted therapy for chromatin abnormalities.

Urothelial carcinoma of the bladder is a major cause of morbidity and mortality worldwide, causing an estimated 150,000 deaths per year[1]. Previous studies have identified multiple regions of somatic copy number alteration, including amplification of *PPARG*, *E2F3*, *EGFR*, *CCND1* and *MDM2*, as well as loss of *CDKN2A* and *RB1* (refs 2, 3). Sequencing of candidate pathways has identified recurrent mutations in *TP53*, *FGFR3*, *PIK3CA*, *TSC1*, *RB1* and *HRAS* (refs 2, 3). Whole-exome sequencing of nine bladder cancers, followed by a replication analysis of 88 cancers, identified mutations at >10% frequency in several chromatin remodelling genes: *KDM6A*, *CREBBP*, *EP300* and *ARID1A* (ref. 4). Focused molecular analyses[5,6] have delineated tumour subtypes and identified kinase-activating *FGFR3* gene fusions[7,8].

We report here a comprehensive, integrated study of 131 high-grade muscle-invasive urothelial bladder carcinomas as part of The Cancer Genome Atlas (TCGA) project. Included are data on DNA copy number, somatic mutation, messenger RNA and microRNA (miRNA) expression, protein and phosphorylated protein expression, DNA methylation, transcript splice variation, gene fusion, viral integration, pathway perturbation, clinical correlates and histopathology to characterize the molecular landscape of urothelial carcinoma. This study identifies a number of mutations and regions of copy number variation that involve genes not previously reported as altered in a significant fraction of bladder cancers. It also identifies potential therapeutic targets in most of the samples analysed.

Demographic, clinical and pathological data

Samples (from 19 tissue source sites) consisted of 131 chemotherapy-naive, muscle-invasive, high-grade urothelial tumours (T2-T4a, Nx, Mx), as well as peripheral blood ($n = 118$) and/or tumour-adjacent, histologically normal-appearing bladder tissue ($n = 23$). Cases were retained only if they met the following criteria: tumour nuclei constituted ≥60% of all nuclei; tumour necrosis was ≤20% of the specimen; and variant histologies (squamous or small cell) were ≤50% (Supplementary Information, section 'Biospecimen collection and clinical data'). Clinical and demographic characteristics are described in Supplementary Data 1.1. Five expert genitourinary pathologists re-reviewed all of the cases for multiple parameters, including the extent of variant histology (Supplementary Fig. 1.1a and Supplementary Information, section 'Biospecimen collection and clinical data').

Somatic DNA alterations

The tumours displayed a large number of DNA alterations, slightly fewer than in lung cancer and melanoma, but more than in other adult malignancies studied by TCGA (Fig. 1)[9]. On average, there were 302 exonic mutations, 204 segmental alterations in genomic copy number and 22 genomic rearrangements per sample. We analysed somatic copy number alterations (CNAs) using both SNP 6.0 arrays and low-pass whole-genome sequencing; the two were strongly concordant (Supplementary Methods 6.1 and Supplementary Fig. 6.1). There were 22 significant arm-level copy number changes (Supplementary Data 6.1.1), and GISTIC (genomic identification of significant targets in cancer) (Supplementary Methods 6.2) identified 27 amplified and 30 deleted recurrent focal somatic CNAs (Supplementary Data 6.2.1 and 6.3.1). Focal amplifications involved genes previously reported to be altered in bladder cancer (Fig. 1c and Supplementary Fig. 6.2.1) and some not previously implicated. The latter included *PVRL4*, *BCL2L1* and *ZNF703*. The most common recurrent focal deletion, seen in 47% of samples, contained *CDKN2A* (9p21.3) and correlated with reduced expression (Fig. 1 and Supplementary Fig. 2.7). Other focal deletions containing <10 genes appeared to target *PDE4D*, *RB1*, *FHIT*, *CREBBP*, *IKZF2*, *FOXQ1*, *FAM190A* (also called *CCSER1*), *LRP1B* and *WWOX*.

Whole-exome sequencing of 130 tumours and matched normal samples targeted 186,260 exons in 18,091 genes (mean coverage 100-fold, with 82% of target bases covered >30×). MuTect[10] identified 39,312 somatic mutations (including 38,012 point mutations and 1,138 indels (insertions or deletions)), yielding mean and median somatic mutation rates of 7.7 and 5.5 per megabase (Mb), respectively (Fig. 1a and Supplementary Table 2.1.1). Thirty-two genes showed statistically significant levels of recurrent somatic mutation (Fig. 1b and Supplementary Table 2.1.2) by analysis using MutSig 1.5 (refs 9, 11) (Supplementary

*A list of authors and affiliations appears at the end of the paper.

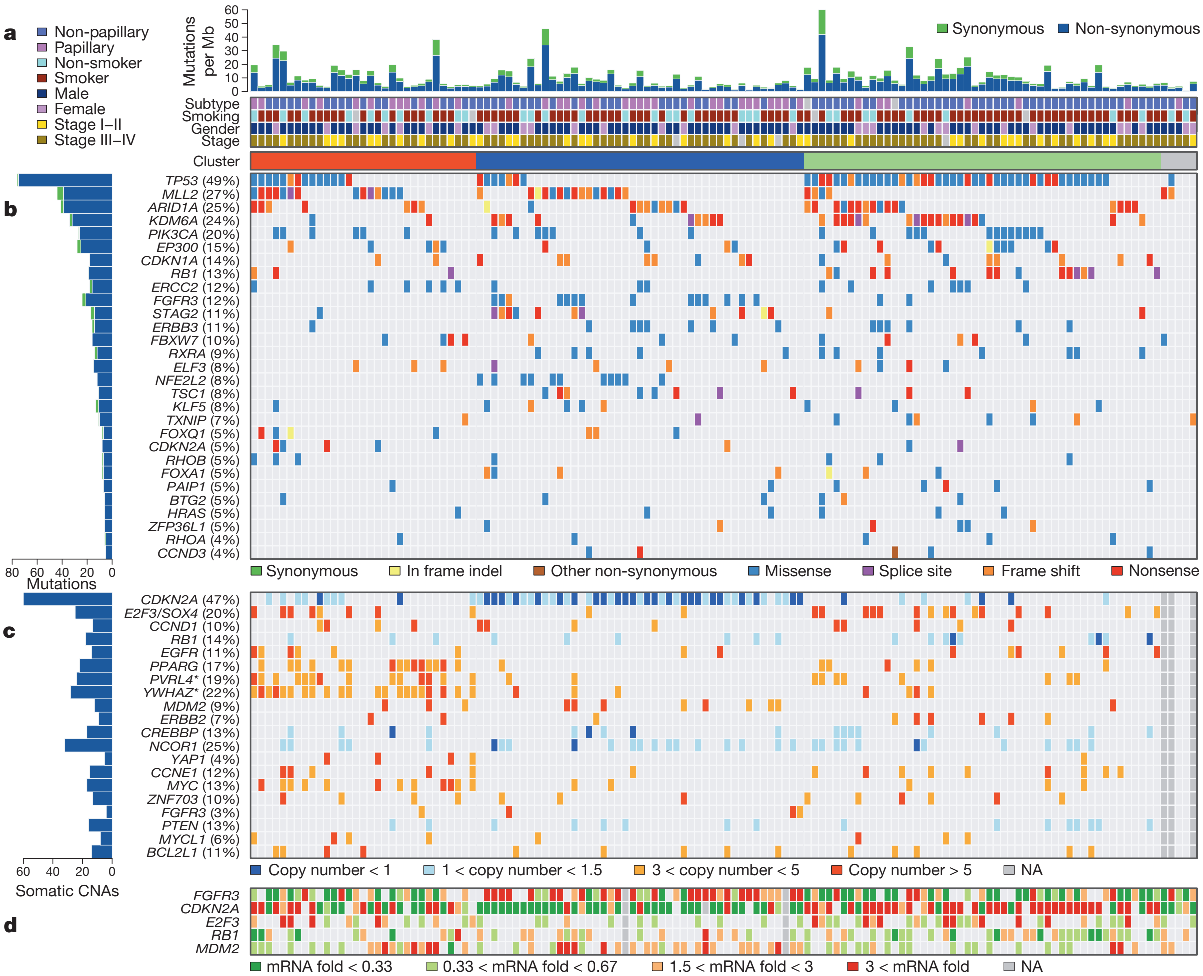


Figure 1 | The genomic landscape of bladder cancer. **a**, Mutation rate and type, histological subtype, smoking status, gender, tumour stage and cluster type. **b**, Genes with statistically significant levels of mutation (MutSig, false discovery rate <0.1) and mutation types. **c**, Deletions and amplifications for genomic regions with statistically significant focal copy number changes (GISTIC2.0). 'Copy number' refers to absolute copy number. Note that two amplification peaks (*) contain several genes, any of which could be the target, as opposed to the single gene listed here. **d**, RNA expression level for selected genes, expressed as fold change from the median value for all samples. Tumour samples were grouped into three clusters (red, blue and green) using consensus NMF clustering (see the main text and Supplementary Fig. 2.1.2). Three samples with no copy number data and two samples with no mutations in the genes were not used in the clustering and are shown in grey.

Methods 2.2). Three other genes identified by MutSig were not considered further because of low or undetectable expression (Supplementary Fig. 2.1.1). A similar analysis considering only mutations in the COSMIC database[2] identified three more significantly mutated genes: *ERBB2*, *ATM* and *CTNNB1* (Supplementary Table 2.1.3). We validated the mutation findings in three ways: targeted re-sequencing of all significantly mutated gene mutations, comparison with RNA-seq data for 123 samples and comparison with whole-genome sequence data for 18 samples. Overall, the validation rate was >99% in selected mutations by a combination of the methods (Supplementary Methods 2.4).

Nearly half (49%) of the samples had *TP53* mutations (Fig. 1b), which were mutually exclusive in their relationship with amplification (9%) and overexpression (29%) of *MDM2*; hence, TP53 function was inactivated in 76% of samples. Most *RB1* mutations were inactivating, were associated with significantly reduced mRNA level (Supplementary Fig. 2.7) and were mutually exclusive with *CDKN2A* deletions (Supplementary Fig. 2.8 and Supplementary Table 2.8.1). *FGFR3* mutations (12%) typically affected known kinase-activating sites. *PIK3CA* mutations were relatively common (20%), clustering in the helical domain near E545 (Supplementary Fig. 2.4). Most *TSC1* mutations (8%) were truncating, and six were homozygous (allele fraction >0.5).

Many of the 32 genes identified in Fig. 1b have not previously been reported as statistically significantly mutated in bladder cancer: *MLL2* (also called *KMT2D*; 27%), *CDKN1A** (14%), *ERCC2** (12%), *STAG2* (11%), *RXRA** (9%), *ELF3** (8%), *NFE2L2* (8%), *KLF5** (8%), *TXNIP* (7%), *FOXQ1** (5%), *RHOB** (5%), *FOXA1* (5%), *PAIP1** (5%), *BTG2** (5%), *ZFP36L1* (5%), *RHOA* (4%) and *CCND3* (4%). The nine genes marked with asterisks have not been reported as significantly mutated genes in any other TCGA cancer type or reported in another study as mutated at >3% frequency[2]. *CDKN1A* (*p21*^{CIP1}), a cyclin-dependent kinase inhibitor[12], had predominantly null or truncating mutations, indicating loss of function. Fifteen of sixteen mutations in *ERCC2*, a nucleotide excision repair gene[13], were deleterious missense mutations, suggesting dominant-negative effects. *ERCC2*-mutant tumours also had significantly fewer C>G mutations than did *ERCC2*-wild-type tumours (Supplementary Figs 2.3.1 and 2.3.2), and they trended towards higher overall mutation rate (Supplementary Fig. 2.12). Seven of twelve mutations in *RXRA* (retinoid X nuclear receptor alpha)[14] occurred at the same amino acid (five S427F; two S427Y) in the ligand-binding domain. Those seven tumours showed increased expression of genes involved in adipogenesis and lipid metabolism (Supplementary Fig. 2.6 and Supplementary Data 2.6.1–2.6.3), suggesting that the mutations cause constitutive activation.

Eleven tumours (8%) had deleterious missense mutations in the Neh2 domain of *NFE2L2*, a transcription factor that regulates the anti-oxidant program in response to oxidative stress[15]. Those tumours

showed markedly increased expression of genes involved in genotoxic metabolism and the reactive oxygen species (ROS) response (Supplementary Figs 2.5.1–2.5.3 and Supplementary Data 2.5.2). Furthermore, nine samples had mutations in redox regulator *TXNIP* (ref. 16) (five of them inactivating) and were mutually exclusive of samples with *NFE2L2* mutations, providing another mechanism for dysregulation of redox metabolism. Predominant inactivating mutations were seen in STAG2, an X-linked cohesin complex component required for separation of sister chromatids during cell division[17] (Supplementary Fig. 2.4).

Unsupervised clustering by non-negative matrix factorization of mutations and focal somatic CNAs in 125 samples identified three distinct groups (Fig. 1a and Supplementary Fig. 2.1.2). Group A (red), classified as 'focally amplified', is highly enriched in focal somatic CNAs in several genes, as well as mutations in *MLL2* (Fig. 1 and Supplementary Tables 2.1.4 and 2.1.5). Group B (blue), classified as 'papillary *CDKN2A*-deficient *FGFR3* mutant', is enriched in papillary histology. Nearly all group B samples show loss of *CDKN2A*, and most have one or more alterations in *FGFR3*. Group C (green), classified as '*TP53*/cell-cycle-mutant', shows *TP53* mutations in nearly all samples, as well as enrichment with *RB1* mutations and amplifications of *E2F3* and *CCNE1* (Fig. 1 and Supplementary Table 2.1.4). These differences in pattern of mutation suggest the possibility of different oncogenic mechanisms.

Seventy-two per cent of the cancers in this study were from current or past smokers, consistent with extensive epidemiological studies indicating an association between smoking and urothelial cancer risk. In contrast with lung cancer, however, there was no statistically significant association between smoking status and the mutational spectrum, frequency of mutation in any significantly mutated gene, occurrence of focal somatic CNAs or expression subtype (Supplementary Tables 2.9.1 and 2.9.2). Never-smokers did have a slightly higher fraction of C>G mutations than did current/former smokers (28.5% versus 23.8%, $P = 0.032$; Supplementary Figs 2.3.2 and 2.3.3). Unsupervised clustering of promoter CpG island DNA methylation data revealed a major subgroup (34%) of tumours (CIMP) characterized by cancer-specific DNA hypermethylation (Supplementary Fig. 7.1). Multivariate regression analysis with age, sex and tumour stage as covariates identified smoking pack-years as the only significant predictor of CIMP phenotype, as has also been reported for colorectal cancer[18].

Fifty-one per cent of mutations overall were Tp*C->(T/G) (Supplementary Table 2.1.1), a class of mutation recently reported to be mediated by one of the DNA cytosine deaminases, APOBEC (refs 19, 20). *APOBEC3B* was expressed at high levels in all of the tumours, suggesting a major role for APOBEC-mediated mutagenesis in bladder carcinogenesis (Supplementary Figs 12.1 and 12.2).

Four genes involved in epigenetic regulation were significantly mutated genes: *MLL2*, *ARID1A*, *KDM6A* and *EP300* (Fig. 1). Truncating mutations were significantly enriched in each of those genes (Supplementary Fig. 2.2 and Supplementary Data 2.2.1–2). Three of the genes had previously been identified as mutated in urothelial cancers[4], but mutation of *MLL2*, which encodes a histone H3 lysine 4 (H3K4) methyltransferase, is a novel finding. Several other chromatin-regulating genes had mutation rates ≥10% but were not statistically significant by MutSig analysis: *MLL3*, *MLL*, *CREBBP*, *CHD7* and *SRCAP*. Many other epigenetic regulators were mutated at lower frequency but were also enriched with truncating mutations, indicating functional significance (Supplementary Fig. 2.2 and Supplementary Data 2.2.1 and 2.2.2). Non-silent mutations in chromatin regulatory genes overall were significantly enriched in bladder cancer in comparison with the entire exome, in contrast with all other epithelial cancers studied so far in the TCGA project (Supplementary Table 2.10). Mutations in *MLL2* and *KDM6A* (the latter encoding a histone H3 lysine 27 (H3K27) demethylase) were mutually exclusive (Supplementary Fig. 2.8 and Supplementary Table 2.8.1), suggesting that mutations in the two genes have redundant downstream effects on carcinogenesis or that the combined loss is synthetically lethal.

Chromosomal rearrangements and viral integration

To identify structural variations and pathogen sequences, we used low-pass, paired-end, whole-genome sequencing (WGS; 6–8× coverage) of 114 tumours and RNA sequencing of all tumours. We detected 2,529 structural aberrations, including 1,153 that involve gene–gene fusions. Among the translocations, 379 were inter-chromosomal, 237 were intra-chromosomal, 274 were the result of inversions and 263 resulted from deletions (Supplementary Table 3.1). We found several recurrent translocations of probable pathogenic significance, including an intra-chromosomal translocation on chromosome 4 involving *FGFR3* and *TACC3* ($n = 3$). The breakpoints were in intron 16 (two cases) or exon 17 (one case) of *FGFR3* and intron 10 of *TACC3* (confirmed by DNA sequencing and RNA-seq). All three lead to fusion mRNA products for which the predicted proteins include the amino-terminal 758 amino acids of FGFR3 fused with the carboxy-terminal 191 amino acids of TACC3 (Fig. 2a). On the basis of the structure of the FGFR3–TACC3 fusion protein, we predict that it can auto-dimerize, leading to constitutive activation of the kinase domain of FGFR3. FGFR3–TACC3 fusion, which was recently described in both glioblastoma[21] and bladder cancer[7,8], represents a promising therapeutic target. The *ERBB2* gene was also involved in translocations in four tumours, all with different fusion partners and all confirmed by DNA sequencing, RNA-seq or both. In one case, exons 4 to 29 of *ERBB2* were fused to the promoter plus exon 1 of *DIP2B*, and the fusion product was amplified (Fig. 2b). Two other fusion products resulted in novel mRNA products, the biological significance of which is not known.

We identified viral DNAs in 7 of 122 tumours (6%), and viral transcripts in 5 of 122 (4%). Three tumours expressed cytomegalovirus (CMV) transcripts (encoding RL5A, RNA2.7, RL9A, RNA1.2, UL5 and UL22A), one expressed BK polyoma virus and one expressed human papilloma virus 16 (HPV16). HPV16 and human herpesvirus 6B DNA were each identified in one other sample but without expression. None of the tumours expressing CMV showed evidence of CMV integration into the host genome, suggesting the presence of a stable episome. In the BK-positive tumour, two BK genes were integrated into *GRB14*, a signalling adaptor protein for receptor tyrosine kinases. In the HPV-16-expressing case, the virus integrated into *BCL2L1*, an apoptosis-regulating gene (Fig. 2c). In that tumour, *BCL2L1* was amplified (~6×) and

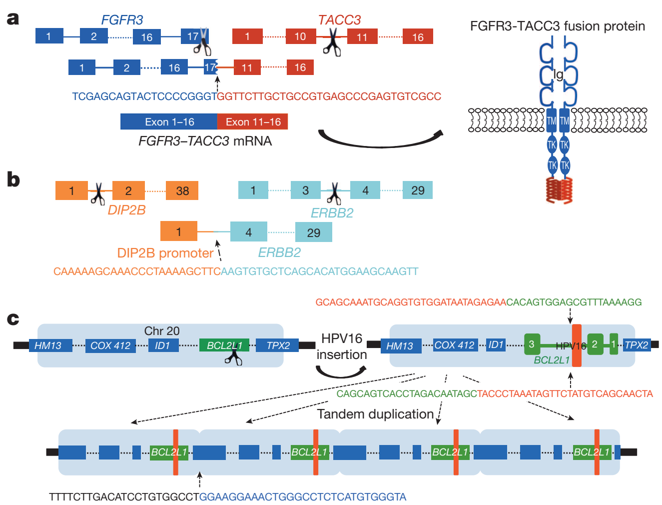


Figure 2 | Structural rearrangements and viral integration. **a**, FGFR3–TACC3 fusion in sample TCGA-CF-A3MH showing the breakpoints in the two genes, the breakpoint junction sequences and the predicted fusion protein. **b**, Rearrangement involving *DIP2B* and *ERBB2* in TCGA-DK-A2I6. The *ERBB2* gene has swapped its promoter with that of *DIP2B*, resulting in overexpression of *ERBB2*. **c**, Insertion of human papilloma virus 16 (HPV16) into the *BCL2L1* gene on chromosome 20 in TCGA-GC-A3I6. The region of *BCL2L1* into which the virus has integrated and the integration junction sequence are shown.

overexpressed (~10× median; >2× any of the other samples). Overall, these findings indicate that viral infection may have a role in the development of a small percentage of urothelial carcinomas.

mRNA, miRNA and protein expression

Analysis of RNA-seq data from 129 tumours identified four clusters (clusters I–IV) (Fig. 3 and Supplementary Fig. 4.1). Cluster I ('papillary-like') is enriched in tumours with papillary morphology ($P = 0.0002$), *FGFR3* mutations ($P = 0.0007$, $q = 0.02$), *FGFR3* copy number gain ($P = 0.04$, $q = 0.1$) and elevated FGFR3 expression ($P < 0.0001$) (Fig. 3a). It includes all three samples with *FGFR3–TACC3* fusions. Cluster I samples also show significantly lower expression of miR-99a and miR-100, miRNAs that downregulate *FGFR3* expression ($P = 0.0002$, Figs 3a and Supplementary Fig. 5.3)[22]. Cluster I samples also show lower expression of miR-145 and miR-125b, which have been reported as frequently downregulated in bladder cancer[23]. Tumours with *FGFR3* alterations, and perhaps other tumours that share the cluster I expression profile, may respond to inhibitors of FGFR or its downstream targets.

Reverse-phase protein array (RPPA) data indicate that clusters I and II express high HER2 (ERBB2) levels and an elevated oestrogen receptor beta (ESR2) signalling signature, indicating potential targets for hormone therapies such as tamoxifen or raloxifene (Fig. 3d). In fact, HER2 protein levels in a subset of the tumours are comparable to those found in TCGA HER2-positive breast cancers[23].

For comparison, we asked whether any of the four clusters show gene signatures similar to those identified in any other tumour type(s) among the first 11 analysed by TCGA. We found that the signature of bladder cancer cluster III ('basal/squamous-like') is similar to that of basal-like breast cancers, as well as squamous cell cancers of the head and neck and lung (Supplementary Fig. 4.2)[24,25]. All four of those cancer types express characteristic epithelial lineage genes, including *KRT14*, *KRT5*, *KRT6A* and *EGFR*. Basal-like subtype[26] and squamous cell subtype[27] of urothelial carcinoma have been independently reported. Many of the samples in bladder cluster III express cytokeratins (that is, *KRT14* and *KRT5*) that were recently reported to mark stem/progenitor cells[26]. Some of those samples also show a level of variant squamous histology (Fig. 3b). Bladder clusters I and II show features similar to those of luminal A breast cancer, with high mRNA and protein expression of luminal breast differentiation markers, including GATA3 and FOXA1 (Fig. 3c). Markers of urothelial differentiation such as the uroplakins (for example, UPK3A) are also highly expressed in clusters I and II, as are the epithelial marker E-cadherin and members of the miR-200 family of miRNAs (which target multiple regulators of epithelial–mesenchymal transition)[28] (Fig. 3c). Taken together, these observations indicate that, despite their diverse tissue origins, some bladder, breast, head and neck and lung cancers share common pathways of tumour development.

To determine whether the expression-based clusters could be seen in other data sets, we used the muscle-invasive bladder cancer samples from ref. 27, hierarchically clustering them with the genes used in our analysis. From the sample dendrogram, we identified four groups (Supplementary Fig. 4.3a). The four groups identified in the data set of ref. 27 correlated well with the four clusters identified in our TCGA data (Supplementary Fig. 4.3b).

When we analysed the RNA-seq data for transcript splice variation using SpliceSeq[29] (Supplementary Information, section 11), one finding of interest was an average of 3% *PKM1* and 97% *PKM2* transcripts in the tumour samples. The *PKM2* isoform of pyruvate kinase is the principal driver of a shift to aerobic glycolysis in tumours (the Warburg effect)[30]. Therefore, urothelial bladder cancers (and other cancer types) may prove sensitive to inhibition of glycolysis or related metabolic pathways.

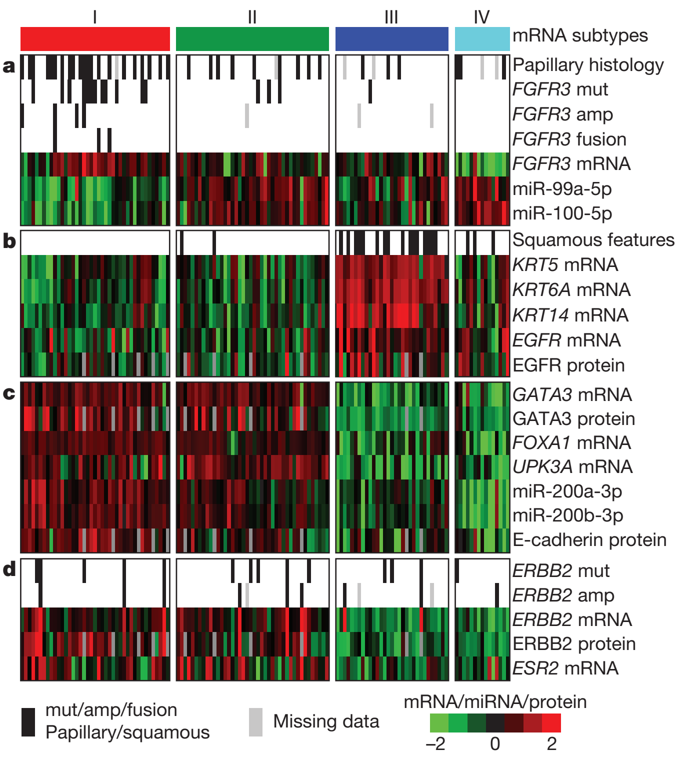


Figure 3 | Expression characteristics of bladder cancer. Integrated analysis of mRNA, miRNA and protein data led to identification of distinct subsets of urothelial carcinoma. Data for mRNA, miRNA and protein were *z*-normalized, and samples were organized in the horizontal direction by mRNA clustering. **a**, Papillary histology, *FGFR3* alterations, *FGFR3* expression and reduced *FGFR3*-related miRNA expression are enriched in cluster I. **b**, Expression of epithelial lineage genes and stem/progenitor cytokeratins are generally high in cluster III, some of which show variant squamous histology. **c**, Luminal breast and urothelial differentiation factors are enriched in clusters I and II. **d**, *ERBB2* mutation and oestrogen receptor beta (*ESR2*) expression are enriched in clusters I and II.

Pathway analysis and therapeutic targeting

Integrated analysis of the mutation and copy-number data revealed three main pathways as frequently dysregulated in bladder cancer: cell cycle regulation (altered in 93% of cases); kinase and phosphatidylinositol-3-OH kinase (PI(3)K) signalling (72%); and chromatin remodelling, including mutations/somatic CNAs in histone-modifying genes (89%) and components of the SWI/SNF nucleosome remodelling complex (64%) (Fig. 4a). To complement these results for well-defined pathways, we applied network analysis methods to examine other possible interactions between genes and pathways (Fig. 4b). In particular, we used the TieDIE algorithm to search for causal regulatory interactions within the PARADIGM network, which connects mutated genes to active transcriptional hubs[31,32]. The analysis identified a sub-network linking mutated histone-modifying genes to a large array of activated transcription factors, indicating potential far-reaching effects of histone modification on other pathways (Supplementary Fig. 8.2.1) converging on MYC/MAX regulation. Both MYC and MAX showed similar levels of pathway activity, independent of mutations in chromatin genes, suggesting that mutations in histone-modifying genes provide just one mechanism for disruption of the MYC/MAX hub. By contrast, tumours with chromatin-related mutations showed differential activity of transcription factors *FOXA2* and *SP1*, implicating de-differentiation processes as a result of the mutations. Our network analysis also identified *HSP90AA1* as a critical signalling hub, indicating that inhibitors of HSP90 may have therapeutic value in urothelial carcinoma. Although the linkages between mutations and transcriptional changes were statistically significant in terms of their proximity in the network (as determined by permutation tests; see Supplementary Fig. 8.2), further studies will be needed to assess the biological relevance of the findings.

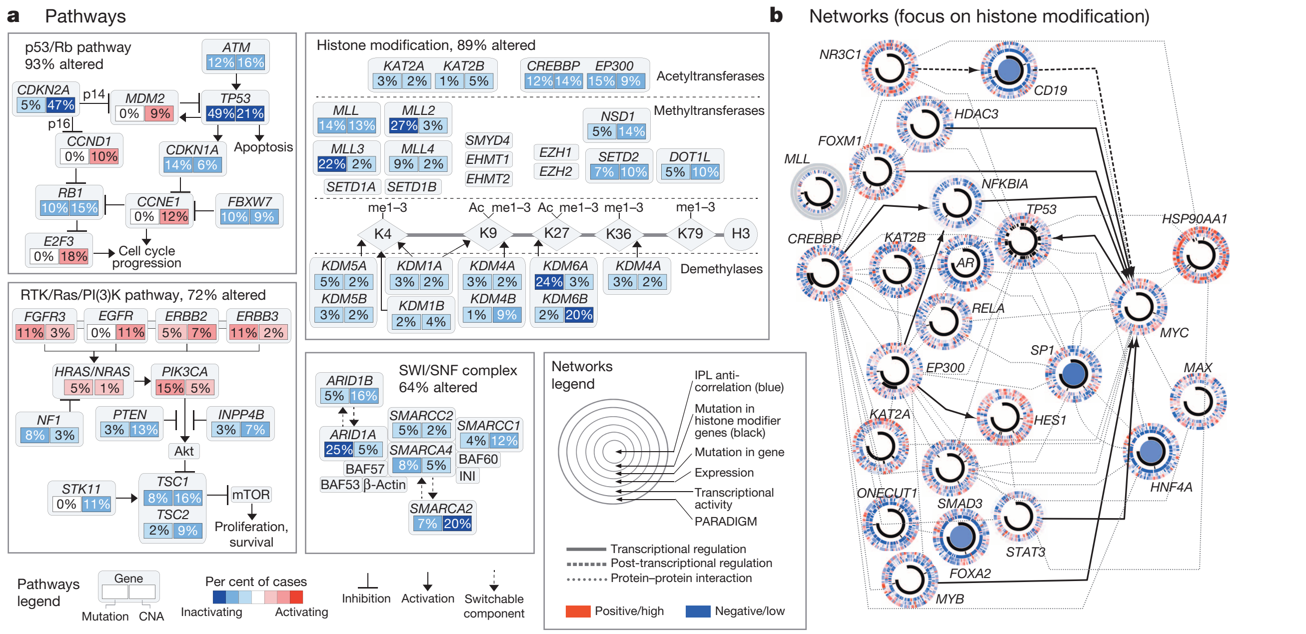


Figure 4 | Altered pathways and networks in bladder cancer. a, Somatic mutations and copy number alterations (CNA) in components of the p53/Rb pathway, RTK/RAS/PI(3)K pathway, histone modification system and SWI/SNF complex. Red, activating genetic alterations; blue, inactivating genetic alterations. Percentages shown denote activation or inactivation of at least one allele. **b**, The network connecting mutated histone-modifying genes to transcription factors with differential activity (methodology and larger implicated network in Supplementary Fig. 8.2.1). Each gene is depicted as a multi-ring circle with various levels of data, plotted such that each 'spoke' in the ring represents a single patient sample (same sample ordering for all genes). 'PARADIGM' ring, bioinformatically inferred levels of gene activity (red, higher activity); 'Transcriptional activity', mean mRNA levels of all of the targets of each transcription factor; 'expression', mRNA levels relative to normal (red, high); 'Mutation in gene', somatic mutation; 'Mutation in histone modifier genes', somatic mutation in at least one such gene; 'IPL anti-correlation', genes with PARADIGM integrated pathway levels (IPLs) inversely correlated with histone-gene mutation status. Gene–gene relationships are inferred using public resources.

Integrated analysis also identified mutations, copy number alterations or RNA expression changes affecting the PI(3)K/AKT/mTOR pathway in 42% of the tumours (Fig. 5a). Included were activating point mutations in *PIK3CA* (17%; potentially responsive to PI(3)K inhibitors), mutation or deletion of *TSC1* or *TSC2* (9%; potentially responsive to mTOR inhibitors) and overexpression of *AKT3* (10%; potentially responsive to AKT inhibitors). We also observed mutations, genomic amplifications or gene fusions that affect the RTK/RAS pathway in 44% of the tumours (Fig. 5b, c). Included were events that can activate FGFR3 (17%; potentially responsive to FGFR inhibitors or antibodies), amplification of EGFR (9%; potentially responsive to EGFR antibodies or inhibitors), mutations of *ERBB3* (6%; potentially

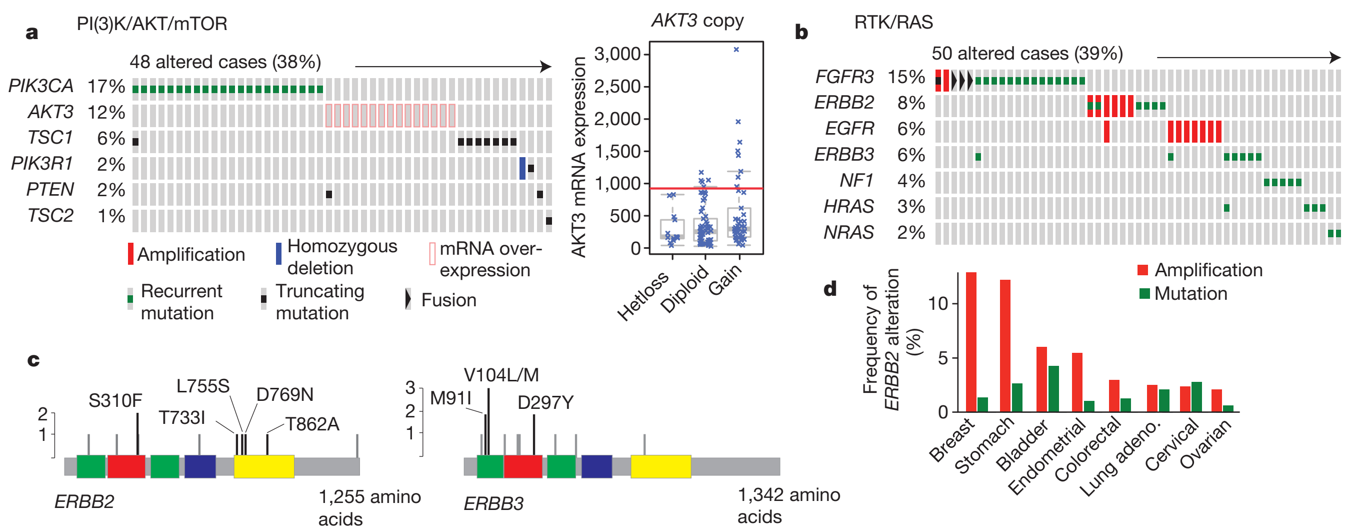


Figure 5 | Potential targets in bladder cancer. a, Alterations in the PI(3)K/AKT/mTOR pathway are mutually exclusive. Tumour samples are shown in columns; genes in rows. Only samples with at least one alteration are shown. AKT3 shows elevated expression in 10% of samples, independent of copy number (right panel). Hetloss, heterozygous loss. **b**, Receptor tyrosine kinases are altered, by any of several different mechanisms (amplification, mutation or fusion), in 45% of samples. Only mutations that are recurrent in this data set or previously reported in COSMIC are shown. **c**, Recurrent mutations in ERBB2 and ERBB3. The mutations shown in black are either recurrent in the TCGA data set or reported in COSMIC. Green, receptor L domain; red, furin-like cysteine-rich region; blue, growth factor receptor domain IV; yellow, tyrosine kinase domain. **d**, *ERBB2* amplifications and recurrent mutations in other cancers profiled by TCGA. Missense mutations were counted in the following positions: G309, S310, L313, R678, T733, L755, V777, D769, V842, T862, R896 and M916I. In-frame insertions were counted between amino acids 774 and 776. Only tumour types with an alteration frequency ≥2% are shown.

sensitive to ERBB kinase inhibitors) and mutation or amplification of *ERBB2* (9%; potentially sensitive to ERBB2 kinase inhibitors or antibodies). *ERBB3* mutations in bladder cancer have been noted previously[4], but statistically significant mutation of *ERBB2* in bladder cancer has not been reported. Both genes are potential therapeutic targets in other diseases[33–35]. Notably, *ERBB2* alterations were approximately as frequent in this study as in TCGA breast cancers, but with fewer amplifications and more mutations (Fig. 5d)[24].

Discussion

This integrated study of 131 invasive urothelial bladder carcinomas provides numerous novel insights into disease biology and delineates multiple potential opportunities for therapeutic intervention. Treatment for muscle-invasive bladder cancer has not advanced beyond cisplatin-based combination chemotherapy and surgery in the past 30 years[36], and no new drugs for the disease have been approved in that time. Median survival for patients with recurrent or metastatic bladder cancer remains 14–15 months with cisplatin-based chemotherapy, and there is no widely recognized second-line therapy[37]. With the exception of a single case report, there is also no known benefit from treatment with newer, targeted agents[38]. Several of the genomic alterations identified in this study, particularly those involving the PI(3)K/AKT/mTOR, CDKN2A/CDK4/CCND1 and RTK/RAS pathways, including ERBB2 (Her-2), ERBB3 and FGFR3, are amenable in principle to therapeutic targeting. Clinical trials based on patients with relevant druggable genomic alterations are warranted.

FGFR3 mutation is a common feature of low-grade non-invasive papillary urothelial bladder cancer, but it occurs at a much lower frequency in high-grade invasive bladder cancer. The cluster analysis in Fig. 3 highlights multiple mechanisms of *FGFR3* activation, and its strong association with papillary morphology. The data presented here suggest a subset of muscle-invasive cancers that can potentially be targeted through *FGFR3*. Similarly, ERBB2 amplification may be targetable by strategies used in breast cancer, by small-molecule tyrosine kinase inhibitors or by novel immunotherapeutic approaches (NCT01353222)[34]. The data here provide further support for several on-going ERBB2-targeted trials in bladder cancer and further define the subpopulation of cancers suited to that approach. Finally, cluster III of the integrated expression profiling analysis reveals the existence of a urothelial carcinoma subtype with cancer stem-cell expression features (including KRT14 and KRT5), perhaps providing another avenue for therapeutic targeting.

The alterations identified in epigenetic pathways also suggest new possibilities for bladder cancer treatment. Ninety-nine (76%) of the tumours analysed here had an inactivating mutation in one or more of the chromatin regulatory genes, and 53 (41%) had at least two such mutations. Overall, the bladder cancers showed a mutational spectrum highly enriched with mutations in chromatin regulatory genes (Supplementary Table 2.10). Furthermore, integrated network analyses revealed a profound impact of those mutations on the activity levels of various transcription factors and pathways implicated in cancer. Drugs that target chromatin modifications—for example, recently developed agents that bind acetyl-lysine binding motifs (bromodomains)—might prove useful for treatment of the subset of bladder tumours that exhibit abnormalities in chromatin-modifying enzymes[39]. Our findings overall indicate bladder cancer as a prime candidate for exploration of that approach to therapy.

METHODS SUMMARY

Tumour and normal samples were obtained with institutional-review-board-approved consent and processed using a modified AllPrep kit (Qiagen) to obtain purified DNA and RNA. Quality-control analyses revealed only modest batch effects (Supplementary Information, section 'Batch effects'). The tumours were profiled using Affymetrix SNP 6.0 microarrays for somatic CNAs, low-pass WGS (HiSeq) for somatic CNAs and translocations, RNA-seq (HiSeq) for mRNA and miRNA expression, Illumina Infinium (HumanMethylation450) arrays for DNA methylation, HiSeq for exome sequencing and RPPA for protein expression and phosphorylation. Statistical analysis and biological interpretation of the data were spearheaded by the TCGA genome data analysis centres. Sequence files are in CGHub (https://cghub.ucsc.edu/). All other molecular, clinical and pathological data are available through the TCGA Data Portal (https://tcga-data.nci.nih.gov/tcga/). The data can be explored through a compendium of next-generation clustered heat maps (http://bioinformatics.mdanderson.org/TCGA/NGCHMPortal/), the cBio Cancer Genomics Portal (http://cbioportal.org), TieDIE (http://sysbiowiki.soe.ucsc.edu/tiedie), SpliceSeq (http://bioinformatics.mdanderson.org/main/SpliceSeq:Overview), MBatch batch effects assessor (http://bioinformatics.mdanderson.org/tcgambatch/) and Regulome Explorer (http://explorer.cancerregulome.org/). Also see Supplementary Information.

Received 17 June; accepted 19 December 2013.

Published online 29 January; corrected online 19 March 2014 (see full-text HTML version for details).

1. Jemal, A. *et al.* Global cancer statistics. *CA Cancer J. Clin.* **61,** 69–90 (2011).
2. Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39,** D945–D950 (2011).
3. Goebell, P. J. & Knowles, M. A. Bladder cancer or bladder cancers? Genetically distinct malignant conditions of the urothelium. *Urol. Oncol.* **28,** 409–428 (2010).
4. Gui, Y. *et al.* Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nature Genet.* **43,** 875–878 (2011).
5. Hurst, C. D., Platt, F. M., Taylor, C. F. & Knowles, M. A. Novel tumor subgroups of urothelial carcinoma of the bladder defined by integrated genomic analysis. *Clin. Cancer Res.* **18,** 5865–5877 (2012).
6. Lindgren, D. *et al.* Integrated genomic and gene expression profiling identifies two major genomic circuits in urothelial carcinoma. *PLoS ONE* **7,** e38863 (2012).
7. Williams, S. V., Hurst, C. D. & Knowles, M. A. Oncogenic FGFR3 gene fusions in bladder cancer. *Hum. Mol. Genet.* **22,** 795–803 (2013).
8. Wu, Y. M. *et al.* Identification of targetable FGFR gene fusions in diverse cancers. *Cancer Discov.* **3,** 636–647 (2013).
9. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).
10. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnol.* **31,** 213–219 (2013).
11. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505,** 495–501 (2014).
12. Warfel, N. A. & El-Deiry, W. S. p21^{WAF1} and tumourigenesis: 20 years after. *Curr. Opin. Oncol.* **25,** 52–58 (2013).
13. Lehmann, A. R. The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. *Genes Dev.* **15,** 15–23 (2001).
14. Tontonoz, P. *et al.* Adipocyte-specific transcription factor ARF6 is a heterodimeric complex of two nuclear hormone receptors, PPARγ and RXRα. *Nucleic Acids Res.* **22,** 5628–5634 (1994).
15. Shibata, T. *et al.* Cancer related mutations in NRF2 impair its recognition by Keap1-Cul3 E3 ligase and promote malignancy. *Proc. Natl Acad. Sci. USA* **105,** 13568–13573 (2008).
16. Zhou, J., Yu, Q. & Chng, W. J. TXNIP (VDUP-1, TBP-2): a major redox regulator commonly suppressed in cancer by epigenetic mechanisms. *Int. J. Biochem. Cell Biol.* **43,** 1668–1673 (2011).
17. Solomon, D. A. *et al.* Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science* **333,** 1039–1043 (2011).
18. Samowitz, W. S. *et al.* Association of smoking, CpG island methylator phenotype, and V600E BRAF mutations in colon cancer. *J. Natl. Cancer Inst.* **98,** 1731–1738 (2006).
19. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149,** 979–993 (2012).
20. Roberts, S. A. *et al.* Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol. Cell* **46,** 424–435 (2012).
21. Singh, D. *et al.* Transforming fusions of FGFR and TACC genes in human glioblastoma. *Science* **337,** 1231–1235 (2012).
22. Oneyama, C. *et al.* MicroRNA-mediated downregulation of mTOR/FGFR3 controls tumor growth induced by Src-related oncogenic pathways. *Oncogene* **30,** 3489–3501 (2011).
23. Yoshino, H. *et al.* Aberrant expression of microRNAs in bladder cancer. *Nature Rev. Urol.* **10,** 396–404 (2013).
24. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70 (2012).
25. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489,** 519–525 (2012).
26. Ho, P. L., Kurtova, A. & Chan, K. S. Normal and neoplastic urothelial stem cells: getting to the root of the problem. *Nature Rev. Urol.* **9,** 583–594 (2012).
27. Sjodahl, G. *et al.* A molecular taxonomy for urothelial carcinoma. *Clin. Cancer Res.* **18,** 3377–3386 (2012).
28. Korpal, M., Lee, E. S., Hu, G. & Kang, Y. The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *J. Biol. Chem.* **283,** 14910–14914 (2008).
29. Ryan, M. C., Cleland, J., Kim, R., Wong, W. C. & Weinstein, J. N. SpliceSeq: a resource for analysis and visualization of RNA-Seq data on alternative splicing and its functional impacts. *Bioinformatics* **28,** 2385–2387 (2012).
30. Christofk, H. R. *et al.* The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* **452,** 230–233 (2008).

31. Vaske, C. J. *et al.* Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* **26,** i237–i245 (2010).
32. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499,** 43–49 (2013).
33. Bose, R. *et al.* Activating HER2 mutations in HER2 gene amplification negative breast cancer. *Cancer Discov.* **3,** 224–237 (2013).
34. Greulich, H. *et al.* Functional analysis of receptor tyrosine kinase mutations in lung cancer identifies oncogenic extracellular domain mutations of ERBB2. *Proc. Natl Acad. Sci. USA* **109,** 14476–14481 (2012).
35. Jaiswal, B. S. *et al.* Oncogenic ERBB3 mutations in human cancers. *Cancer Cell* **23,** 603–617 (2013).
36. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology for Bladder Cancer. Vol. **1.2012,** http://www.nccn.org/professionals/physician_gls/f_guidelines.asp#site (2012).
37. von der Maase, H. *et al.* Long-term survival results of a randomized trial comparing gemcitabine plus cisplatin, with methotrexate, vinblastine, doxorubicin, plus cisplatin in patients with bladder cancer. *J. Clin. Oncol.* **23,** 4602–4608 (2005).
38. Iyer, G. *et al.* Genome sequencing identifies a basis for everolimus sensitivity. *Science* **338,** 221 (2012).
39. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468,** 1067–1073 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to all of the patients and families who contributed to this study, as well as C. Gunter and L. Chastain for scientific editing and M. Sheth, J. Zhang and C. Ron Bouchard for administrative support. This work was supported by the following grants from the United States National Institutes of Health: U54 HG003273, U54 HG003067, U54 HG003079, U24 CA143799, U24 CA143835, U24 CA143840, U24 CA143843, U24 CA143845, U24 CA143848, U24 CA143858, U24 CA143866, U24 CA143867, U24 CA143882, U24 CA143883, U24 CA144025 and P01 CA120964. Additional personnel and funding sources are acknowledged in the Supplementary Information.

Author Information The primary and processed data used to generate the analyses presented here can be downloaded by registered users from The Cancer Genome Atlas at https://tcga-data.nci.nih.gov/tcga/tcgaDownload.jsp. All of the primary sequence files are deposited in CGHub and all other data are deposited at the Data Coordinating Center (DCC) for public access (http://cancergenome.nih.gov/, https://cghub.ucsc.edu/ and https://tcga-data.nci.nih.gov/docs/publications/blca_2013/). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.N.W. (jweinste@mdanderson.org), S.P.L. (slerner@bcm.edu) or D.J.K. (dk@rics.bwh.harvard.edu).

Author Contributions The Cancer Genome Atlas research network contributed collectively to this study. Biospecimens were provided by the tissue source sites and processed by the Biospecimen Core Resource. Data generation and analyses were performed by the genome-sequencing centres, cancer genome-characterization centres and genome data analysis centres. All data were released through the Data Coordinating Center. Project activities were coordinated by the NCI and NHGRI project teams. We also acknowledge the following TCGA investigators of the Bladder Analysis Working Group who contributed substantially to the project. Project leaders: J. N. Weinstein and S. P. Lerner. Data coordinator: C. J. Creighton. Analysis coordinators: R. Akbani and J. Kim. Manuscript coordinator: M. B. Morgan. Project coordinator: M. Sheth. Writing team: J. N. Weinstein, D. J. Kwiatkowski, S. P. Lerner, C. J. Creighton, P. W. Laird, R. Kucherlapati, R. Akbani, X. Su, K. A. Hoadley and M. C. Ryan. Clinical expertise: S. Lerner, D. J. Kwiatkowski, J. E. Rosenberg and D. Bajorin. Pathology review: H. Al-Ahmadie, B. A. Czerniak, D. Hansel, V. Reuter and B. Robinson. DNA sequence and copy number analysis: J. Kim, D. J. Kwiatkowski, A. D. Cherniack and J. E. Rosenberg. DNA methylation analysis: P. W. Laird and T. Hinoue. mRNA analysis: K. A. Hoadley, W. Y. Kim, J. S. Damrauer, W. Zhang, Y. Liu and R. Akbani. miRNA analysis: G. Robertson and A. J. Mungall. Transcript splicing analysis: M. Ryan and J. N. Weinstein. Protein analysis: R. Akbani and G. B. Mills. APOBEC: D. A. Gordenin. Pathway/integrated analysis: C. J. Creighton, N. Schultz, Evan O. Paull and J. Stuart. Chromosomal rearrangements and viral integration: X. Su, R. Kucherlapati, N. Santoso, S. Lee and M. Parfenov. Batch effects: R. Akbani and J. N. Weinstein. Manuscript review: R. Gibbs, C. Gunter and M. Meyerson. Contact PIs: J. N. Weinstein, S. P. Lerner and D. J. Kwiatkowski.

The Cancer Genome Atlas Research Network

Analysis working group: The University of Texas MD Anderson Cancer Center John N. Weinstein[1,2], Rehan Akbani[1], Bradley M. Broom[1], Wenyi Wang[1], Roeland G. W. Verhaak[1], David McConkey[3]; **Baylor College of Medicine** Seth Lerner[4,5], Margaret Morgan[5,6], Chad J. Creighton[7], Carolyn Smith[8]; **Broad Institute** David J. Kwiatkowski[9,10,11], Andrew D. Cherniack[9], Jaegil Kim[9], Chandra Sekhar Pedamallu[9,12], Michael S. Noble[9]; **Memorial Sloan-Kettering Cancer Center** Hikmat A. Al-Ahmadie[13], Victor E. Reuter[13], Jonathan E. Rosenberg[13], Dean F. Bajorin[13], Bernard H. Bochner[13], David B. Solit[13]; **Oregon Health and Science University, Department of Urology** Theresa Koppie[14]; **Weill Medical College of Cornell University** Brian Robinson[15]; **National Institute of Environmental Health Sciences** Dmitry A. Gordenin[16], David Fargo[16], Leszek J. Klimczak[16], Steven A. Roberts[16]; **Optimum Therapeutics LLC** Jessie Au[17]; **University of Southern California Epigenome Center** Peter W. Laird[18], Toshinori Hinoue[18]; **Computational Biology Center, Memorial Sloan-Kettering Cancer Center** Nikolaus Schultz[19], Ricardo Ramirez[19]; **UCSD Department of Pathology** Donna Hansel[20]; **Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill** Katherine A. Hoadley[21], William Y. Kim[21,22,23]; **Department of Genetics, University of North Carolina at Chapel Hill** Jeffrey S. Damrauer[21,22]; **The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University** Stephen B. Baylin[24]; **Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency** Andrew J. Mungall[30], A. Gordon Robertson[30], Andy Chu[30].

Genome Sequencing Center: Broad Institute David J. Kwiatkowski[9,10,11], Carrie Sougnez[9], Kristian Cibulskis[9], Lee Lichtenstein[9], Andrey Sivachenko[9], Chip Stewart[9], Michael S. Lawrence[9], Gad Getz[9,25], Eric Lander[9], Stacey B. Gabriel[9].

Genome characterization centres: Dan L. Duncan Cancer Center, Human Genome Sequencing Center, Baylor College of Medicine Chad J. Creighton[7], Lawrence Donehower[7,26]; **Broad Institute** Andrew D. Cherniack[9], Jaegil Kim[9], Scott L. Carter[9], Gordon Saksena[9], Steven E. Schumacher[9,27], Carrie Sougnez[9], Samuel S. Freeman[9], Joonil Jung[9], Chandra Sekhar Pedamallu[9,12], Ami S. Bhatt[9,12], Trevor Pugh[9,12], Gad Getz[9,25], Rameen Beroukhim[9,12,28], Stacey B. Gabriel[9], Matthew Meyerson[9,12,29]; **Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency** Andrew J. Mungall[30], A. Gordon Robertson[30], Andy Chu[30], Adrian Ally[30], Miruna Balasundaram[30], Yaron S. N. Butterfield[30], Noreen Dhalla[30], Carrie Hirst[30], Robert A. Holt[30], Steven J. M. Jones[30], Darlene Lee[30], Haiyan I. Li[30], Marco A. Marra[30], Michael Mayo[30], Richard A. Moore[30], Jacqueline E. Schein[30], Payal Sipahimalani[30], Angela Tam[30], Nina Thiessen[30], Tina Wong[30], Natasja Wye[30], Reanne Bowlby[30], Eric Chuah[30], Ranabir Guin[30], Steven J. M. Jones[30], Marco A. Marra[30]; **University of Southern California Epigenome Center** Toshinori Hinoue[18], Hui Shen[18], Moiz S. Bootwalla[18], Timothy Triche Jr[18], Phillip H. Lai[18], David J. Van Den Berg[18], Daniel J. Weisenberger[18], Peter W. Laird[18]; **UCSD Department of Pathology** Donna Hansel[20]; **Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill** Katherine A. Hoadley[21], Saianand Balu[21], Tom Bodenheimer[21], Jeffrey S. Damrauer[21,22] Alan P. Hoyle[21], Stuart R. Jefferys[21], Shaowu Meng[21], Lisle E. Mose[21], Janae V. Simons[21], Mathew G. Soloway[21], Junyuan Wu[21], William Y. Kim[21,22,23], Joel S. Parker[21,22], D. Neil Hayes[21,31]; **Research Computing Center, University of North Carolina at Chapel Hill** Jeffrey Roach[32]; **Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill** Elizabeth Buda[33]; **Department of Biology, University of North Carolina at Chapel Hill** Corbin D. Jones[33,34], Piotr A. Mieczkowski[34], Donghui Tan[34], Umadevi Veluvolu[34], Scot Waring[34]; **Eshelman School of Pharmacy, University of North Carolina at Chapel Hill** J. Todd Auman[35]; **Department of Genetics, University of North Carolina at Chapel Hill** Charles M. Perou[22], Matthew D. Wilkerson[22]; **Department of Genetics, Harvard Medical School** Netty Santoso[36], Michael Parfenov[36], Xiaojia Ren[36], Angeliki Pantazi[36], Angela Hadjipanayis[36,37], Jonathan Seidman[36], Raju Kucherlapati[36,37]; **The Center for Biomedical Informatics, Harvard Medical School** Semin Lee[38], Lixing Yang[38], Peter J. Park[37,38,39]; **Cancer Biology Division, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University** Stephen B. Baylin[24]; **Division of Genetics, Brigham and Women's Hospital** Andrew Wei Xu[37]; **Institute for Applied Cancer Science, Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center** Alexei Protopopov[40], Jianhua Zhang[40], Christopher Bristow[40], Harshad S. Mahadeshwar[40], Sahil Seth[40], Xingzhi Song[40], Jiabin Tang[40], Dong Zeng[40], Lynda Chin[9,40]; **The University of Texas MD Anderson Cancer Center, Department of Pathology** Charles Guo[41]

Genome data analysis centres: The University of Texas M.D. Anderson Cancer Center John N. Weinstein[1,2], Rehan Akbani[1], Bradley M. Broom[1], David McConkey[3], Tod D. Casasent[10], Wenbin Liu[1,2], Zhenlin Ju[1,2], Thomas Motter[1], Bo Peng[1], Michael Ryan[1], Wenyi Wang[1], Roeland G. W. Verhaak[1], Xiaoping Su[1], Ji-Yeon Yang[1,2], Philip L. Lorenzi[1], Hui Yao[1], Nianxiang Zhang[1], Jiexin Zhang[1], Gordon B. Mills[2]; **Broad Institute** Jaegil Kim[9], Michael S. Noble[9], Juok Cho[9], Daniel DiCara[9], Scott Frazer[9], Nils Gehlenborg[9], David I. Heiman[9], Pei Lin[9], Yingchun Liu[9], Petar Stojanov[9,12], Doug Voet[9], Hailei Zhang[9], Lihua Zou[9], Lynda Chin[9,40], Gad Getz[9,25]; **Institute for Systems Biology** Brady Bernard[42], Dick Kreisberg[42], Sheila Reynolds[42], Hector Rovira[42], Ilya Shmulevich[42]; **Computational Biology Center, Memorial Sloan-Kettering Cancer Center** Ricardo Ramirez[19], Nikolaus Schultz[19], Jianjiong Gao[19], Anders Jacobsen[19], B. Arman Aksoy[19], Yevgeniy Antipin[19], Giovanni Ciriello[19], Gideon Dresdner[19], Benjamin Gross[19], William Lee[19], Boris Reva[19], Ronglai Shen[19], Rileen Sinha[19], S. Onur Sumer[19], Nils Weinhold[19], Marc Ladanyi[19], Chris Sander[19]; **Buck Institute for Research on Aging** Christopher Benz[43]; **University of California Santa Cruz** Daniel Carlin[44], David Haussler[44], Sam Ng[44], Evan O. Paull[44], Joshua Stuart[44], Jing Zhu[44]; **Department of Pathology, MD Anderson Cancer Center** Yuexin Liu[45], Wei Zhang[45]; **Helen Diller Family Comprehensive Cancer Center, University of California** Barry S. Taylor[46]

Biospecimen core resource: The Research Institute at Nationwide Children's Hospital Tara M. Lichtenberg[47], Erik Zmuda[47], Thomas Barr[47], Aaron D. Black[47], Myra George[47], Benjamin Hanf[47], Carmen Helsel[47], Cynthia McAllister[47], Nilsa C. Ramirez[47,48], Teresa R. Tabler[47], Stephanie Weaver[47], Lisa Wise[47], Jay Bowen[47], Julie M. Gastier-Foster[47,48]

Tissue source sites: The University of Texas MD Anderson Cancer Center John N. Weinstein[1,2]; **Scott Department of Urology, Baylor College of Medicine** Seth Lerner[4,5], Weiguo Jian[4,5], Sebrina Tello[4,5]; **Texas Cancer Research Biobank (TCRB), Baylor College of Medicine** Michael Ittman[5,49], Patricia Castro[5,49], Whitney D. McClenden[5], Margaret Morgan[5,6], Richard Gibbs[5,6]; **Broad Institute** Yingchun Liu[9]; **Analytical Biological Services, Inc.** Charles Saller[50], Katherine Tarvin[50]; **Cleveland Clinic Foundation** Jennifer M. DiPiero[51], Jennifer Owens[51]; **Georgia Regents University Cancer Center** Roni Bollag[52], Qiang Li[52], Paul Weinberger[52]; **Helen F. Graham Cancer Center at Christiana Care** Christine Czerwinski[53], Lori Huelsenbeck-Dill[53], Mary Iacocca[53], Nicholas Petrelli[53], Brenda Rabeno[53], Pat Swanson[53]; **International Genomics Consortium** Troy Shelton[54], Erin Curley[54], Johanna Gardner[54], David Mallery[54], Robert Penny[54]; **ILSbio, LLC** Nguyen Van Bang[55,56], Phan Thi Hanh[55,56], Bernard Kohl[55], Xuan Van Le[55], Bui Duc Phu[55,56], Richard Thorp[55], Nguyen Viet Tien[55,56], Le Quang Vinh[55,56]; **IU School of Medicine** George Sandusky[57]; **Lahey Hospital and Medical Center** Eric Burks[58], Kimberly Christ[58], Jason Gee[58], Antonia Holway[58], Alireza Moinzadeh[58], Andrea Sorcini[58], Travis Sullivan[58]; **Memorial Sloan-Kettering Cancer Center** Hikmat A. Al-Ahmadie[13], Dean F. Bajorin[13], Bernard H. Bochner[13], Ilana R. Garcia-Grossman[13], Ashley M. Regazzi[13], David B. Solit[13], Jonathan E. Rosenberg[13], Victor E. Reuter[13]; **Oregon Health and Science University, Department of Urology** Theresa Koppie[14]; **University of North Carolina, Lineberger Cancer Center** Lori Boice[59], Wendy Kimryn Rathmell[59], Leigh Thorne[59]; **University of Pittsburgh** Sheldon Bastacky[60], Benjamin Davies[60], Rajiv Dhir[60], Jeffrey Gingrich[60], Ronald Hrebinko[60], Jodi Maranchie[60], Joel Nelson[60], Anil Parwani[60]; **Roswell Park Cancer Institute** Wiam Bshara[61], Carmelo Gaudioso[61], Carl Morrison[61]; **Ontario Tumour Bank—Hamilton site, St Joseph's Healthcare Hamilton** Vina Alexopoulou[62], John Bartlett[62], Jay Engel[62], Sugy Kodeeswaran[62]; **The University of Chicago** Tatjana Antic[63], Peter H. O'Donnell[63], Norm D. Smith[63], Gary D. Steinberg[63]; **University of Miami, Sylvester Comprehensive Cancer Center** Sophie Egea[64], Carmen Gomez-Fernandez[64], Lynn Herbert[64], Merce Jorda[64], Mark Soloway[64]; **UT Southwestern Medical Center** Allison Beaver[65], Suzie Carter[65], Payal Kapur[65], Cheryl Lewis[65], Yair Lotan[65]; **Weill Medical College of Cornell University** Brian Robinson[15]; **UCSD Department of Pathology** Donna Hansel[20]; **The University of Texas MD Anderson Cancer Center, Department of Pathology** Charles Guo[41], Jolanta Bondaruk[41], Bogdan Czerniak[41]

Disease working group: The University of Texas MD Anderson Cancer Center Rehan Akbani[1], Bradley M. Broom[1], Yuexin Liu[45], Wei Zhang[45], John N. Weinstein[1,2]; **Scott Department of Urology, Baylor College of Medicine** Seth Lerner[4,5]; **Baylor College of Medicine** Margaret Morgan[5,6]; **Broad Institute** Jaegil Kim[9], Andrew D. Cherniack[9], Samuel S. Freeman[9], Chandra Sekhar Pedamallu[9,12], Michael S. Noble[9], David J. Kwiatkowski[9,10,11]; **Memorial Sloan-Kettering Cancer Center** Hikmat A. Al-Ahmadie[13], Dean F. Bajorin[13], Bernard H. Bochner[13], David B. Solit[13], Jonathan E. Rosenberg[13], Victor E. Reuter[13]; **Oregon Health and Science University, Department of Urology** Theresa Koppie[14]; **Weill Medical College of Cornell University** Brian Robinson[15]; **Stanford University, Department of Urology** Eila Skinner[66]; **Computational Biology Center, Memorial Sloan-Kettering Cancer Center** Ricardo Ramirez[19], Nikolaus Schultz[19]; **UCSD Department of Pathology** Donna Hansel[20]; **Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill** William Y. Kim[21,22,23]; **The University of Texas MD Anderson Cancer Center, Department of Pathology** Charles Guo[41], Jolanta Bondaruk[41], Kenneth Aldape[41], Bogdan Czerniak[41]

Data coordination centre: SRA International Mark A. Jensen[67], Ari B. Kahn[67], Todd D. Pihl[67], David A. Pot[67], Deepak Srinivasan[67], Yunhu Wan[67]

Project team: MLF Consulting Martin L. Ferguson[68]; **National Cancer Institute** Jean Claude Zenklusen[69], Tanja Davidsen[69], John A. Demchok[69], Kenna R. Mills Shaw[3,69], Margi Sheth[69], Roy Tarnuzzer[69], Zhining Wang[69], Liming Yang[69]; **National Human Genome Research Institute** Carolyn Hutter[70], Bradley A. Ozenberger[70], Heidi J. Sofia[70]; **Scimentis, LLC** Greg Eley[71]

[1]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. [2]Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. [3]The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. [4]Scott Department of Urology, Baylor College of Medicine, Houston, Texas 77030, USA. [5]Texas Cancer Research Biobank (TCRB), Baylor College of Medicine, Houston, Texas 77030, USA. [6]Human Genome Sequencing Center at Baylor College of Medicine, Houston, Texas 77030, USA. [7]Dan L. Duncan Cancer Center, Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA. [8]Baylor College of Medicine, Houston, Texas 77030, USA. [9]The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University Cambridge, Massachusetts 02142, USA. [10]Brigham and Women's Hospital, 75 Francis St, Boston, Massachusetts 02115, USA. [11]Harvard Medical School, Boston, Massachusetts 02115, USA. [12]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. [13]Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. [14]Oregon Health and Science University, Department of Urology, 3303 SW Bond Avenue, CHH10U, Portland, Oregon 97239, USA. [15]Weill Medical College of Cornell University, New York, New York 10065, USA. [16]National Institute of Environmental Health Sciences, 111 T.W. Alexander Drive, Research Triangle Park, North Carolina 27709, USA. [17]Optimum Therapeutics LLC, 9363 Towne Centre Drive, San Diego, California 92121, USA. [18]University of Southern California Epigenome Center, University of Southern California, Los Angeles, California 90033, USA. [19]Computational Biology Center, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. [20]UCSD Department of Pathology 9500 Gilman Drive, La Jolla, California 92093, USA. [21]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [22]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [23]Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [24]Cancer Biology Division, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore, Maryland 21231, USA. [25]Massachusetts General Hospital, Cancer Center and Department of Pathology, 55 Fruit Street, Boston, Massachusetts 02114, USA. [26]Department of Molecular Virology and Microbiology, Baylor College of Medicine, 1 Baylor Plaza, Houston, Texas 77030, USA. [27]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. [28]Department of Medicine, Harvard Medical School, Boston, Massachusetts 02215, USA. [29]Department of Pathology, Harvard Medical School, Boston, Massachusetts 02215, USA. [30]Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia V5Z 4S6, Canada. [31]Department of Internal Medicine, Division of Medical Oncology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [32]Research Computing Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [33]Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [34]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [35]Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [36]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [37]Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. [38]The Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts 02115, USA. [39]Informatics Program, Children's Hospital, Boston, Massachusetts 02115, USA. [40]Institute for Applied Cancer Science, Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. [41]The University of Texas MD Anderson Cancer Center, Department of Pathology, Unit 085; 1515 Holcombe Boulevard, Houston, Texas 77030, USA. [42]Institute for Systems Biology, 401 Terry Ave N, Seattle, Washington 98109, USA. [43]Buck Institute for Research on Aging; 8001 Redwood Blvd, Novato, California 94945, USA. [44]University California Santa Cruz, 1156 High Street, Santa Cruz, California 95064, USA. [45]Department of Pathology, MD Anderson Cancer Center, Houston, Texas 77030, USA. [46]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, California 94158, USA. [47]The Research Institute at Nationwide Children's Hospital, Columbus, Ohio 43205, USA. [48]The Ohio State University, Columbus, Ohio 43210, USA. [49]Department of Pathology, Baylor College of Medicine, Houston, Texas 77030, USA. [50]Analytical Biological Services, Inc., 701 Cornell Drive, Wilmington, Delaware 19801, USA. [51]Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, Ohio 44195, USA. [52]Georgia Regents University Cancer Center. Augusta, Georgia 30912, USA. [53]Helen F. Graham Cancer Center at Christiana Care, 4701 Ogletown Stanton Road, Newark, Delaware 19713, USA. [54]International Genomics Consortium, 445 N. Fifth Street, Phoenix, Arizona 85004, USA. [55]ILSbio, LLC 100 Radcliffe Drive, Chestertown, Maryland 21620, USA. [56]Hue Central Hospital, Hue City, Vietnam. [57]IU School of Medicine, Med Science Bldg, Room 128A, 635 Barnhill Drive, Indianapolis, Indiana 46202, USA. [58]Lahey Hospital and Medical Center, Burlington, Massachusetts 01805, USA. [59]University of North Carolina, Lineberger Cancer Center, 450 West Drive, Chapel Hill, North Carolina 27599, USA. [60]University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA. [61]Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, New York 14063, USA. [62]Ontario Tumour Bank—Hamilton site, St Joseph's Healthcare Hamilton, Hamilton, Ontario L8N 3Z5, Canada. [63]The University of Chicago, Chicago, Illinois 60637, USA. [64]University of Miami, Sylvester Comprehensive Cancer Center, 1550 NW 10th Avenue, Miami, Florida 33136, USA. [65]UT Southwestern Medical Center 5323 Harry Hines Blvd, Dallas, Texas 75390-9110, USA. [66]Stanford University, Department of Urology, 300 Pasteur Drive, Suite S287, Stanford, California 94305, USA. [67]SRA International, Fairfax, Virginia 22033, USA. [68]MLF Consulting, Arlington, Massachusetts 02474, USA. [69]National Cancer Institute, 31 Center Drive, 3A20, Bethesda, Maryland 20892, USA. [70]National Human Genome Research Institute, 5635 Fishers Lane, Rockville, Maryland 20852, USA. [71]Scimentis, LLC, Atlanta, Georgia 30666, USA.

ARTICLE

doi:10.1038/nature13145

Coupling of angiogenesis and osteogenesis by a specific vessel subtype in bone

Anjali P. Kusumbe[1]*, Saravana K. Ramasamy[1]* & Ralf H. Adams[1,2]

The mammalian skeletal system harbours a hierarchical system of mesenchymal stem cells, osteoprogenitors and osteoblasts sustaining lifelong bone formation. Osteogenesis is indispensable for the homeostatic renewal of bone as well as regenerative fracture healing, but these processes frequently decline in ageing organisms, leading to loss of bone mass and increased fracture incidence. Evidence indicates that the growth of blood vessels in bone and osteogenesis are coupled, but relatively little is known about the underlying cellular and molecular mechanisms. Here we identify a new capillary subtype in the murine skeletal system with distinct morphological, molecular and functional properties. These vessels are found in specific locations, mediate growth of the bone vasculature, generate distinct metabolic and molecular microenvironments, maintain perivascular osteoprogenitors and couple angiogenesis to osteogenesis. The abundance of these vessels and associated osteoprogenitors was strongly reduced in bone from aged animals, and pharmacological reversal of this decline allowed the restoration of bone mass.

Blood vessels mediate the transport of circulating cells, oxygen, nutrients and waste products, but also provide so-called angiocrine signals controlling organ growth and homeostasis[1–4]. In addition to well-known differences between arteries, capillaries and veins, endothelial cells (ECs) are highly heterogenous and acquire specialized functional properties in local microenvironments[5,6]. The nature of these specialization processes and the underlying regulation remain largely unknown.

In the mammalian skeletal system, growth of the vascular network is regulated by signals provided by chondrocytes and other bone cells, among which the vascular endothelial growth factor (VEGF) is best understood[7,8]. Conversely, blood vessels are thought to influence the osteogenic generation of new bone[9,10]. In addition to earlier studies highlighting the close proximity of vascular and osteoblastic cells, potential roles of angiogenic blood vessel growth in fracture healing have been proposed[11,12]. It also has been suggested that alterations in the skeletal microvasculature might be linked to compromised haematopoiesis and osteogenesis in human subjects with primary osteoporosis or during old age[13,14]. However, direct evidence for such disease-causing or age-related alterations is lacking and our understanding of the normal organization, functional specialization and precise function of the skeletal vasculature is incomplete.

Vessel architecture and bone oxygenation

Previous work has shown that molecular and structural differences distinguish arteries and distal arterioles in bone from sinusoidal capillaries[15–18]. Sinusoidal capillaries form highly branched and irregular networks, and are primarily found between haematopoietic cells within the marrow cavity. Arteries contain comparably few side branches and, as earlier publications[19,20] have shown, preferentially enter long bone through the femoral or tibial head. The precise overall organization of the skeletal vasculature has remained poorly understood because of technical difficulties associated with the processing of bone combined with the loss of crucial three-dimensional information in thin tissue sections. Revised immunohistochemistry protocols (see Methods) have now allowed us to image the bone vasculature at high resolution. To characterize the organization of arteries, α-smooth muscle actin (α-SMA) staining on thick tibia sections (300 μm) was performed. As proposed previously[19,20], straight and smooth muscle cell-covered arteries entered through the bone cortex (Fig. 1a). The central diaphysis contained few, largely unbranched arteries (diameter 16–27 μm). In contrast, substantial branching was seen in the metaphysis when arteries (diameter 5–15 μm) approached the growth plate (Fig. 1a), a structure that in small rodents, unlike in humans, persists into adulthood. Some distal arterioles terminated at capillaries in the endosteum (the connective tissue lining the inner surface of compact bone), whereas most termination points were found in the metaphysis near the growth plate (Fig. 1b, c and Extended Data Fig. 1a). Veins were located in the central diaphysis, connected to metaphyseal capillaries and exited through the cortical bone collar (Extended Data Fig. 2a). Thus, arterial oxygen and nutrient-rich blood is preferentially fed into capillaries near the growth plate, which might affect the local metabolic status. Analysis of tibia sections by pimonidazole staining indicated low oxygenation of the diaphysis whereas the metaphysis was not hypoxic (Extended Data Fig. 2b–e). Likewise, hypoxia-inducible factor 1 α (HIF1-α) immunostaining abundantly labelled cell nuclei throughout the diaphysis and in secondary ossification centres but not in metaphysis (Extended Data Fig. 3a). The plasma membrane lactate transporter MCT4 (also known as Slc16a3), a hypoxia-regulated protein[21], and glucose transporter-1 (Glut1), which is induced by HIF signalling and low glucose[22], were abundant throughout the diaphysis but not in the metaphysis (Extended Data Fig. 3b–d). Conversely, phosphorylated mitogen-activated protein kinase (phospho-Erk1/2), a marker of cell growth and high metabolic activity[23], was strongly enriched in metaphyseal cells (Extended Data Fig. 3e). These data indicate that the spatial distribution of arteries and arteriole–capillary connections determine regional differences in oxygenation and metabolic activity.

Identification of a distinct EC subpopulation in bone

In addition to revised immunofluorescence protocols, we visualized bone vessels by endothelial-cell-specific expression of green fluorescent protein (GFP) in tamoxifen-treated *Cdh5(PAC)-CreERT2*, *Rosa26-mT/mG* double transgenic mice[24,25]. Imaging of the bone microvasculature with both approaches uncovered structurally distinct capillary subsets. Consistent

[1]Max Planck Institute for Molecular Biomedicine, Department of Tissue Morphogenesis, D-48149 Münster, Germany. [2]University of Münster, Faculty of Medicine, D-48149 Münster, Germany.
*These authors contributed equally to this work.

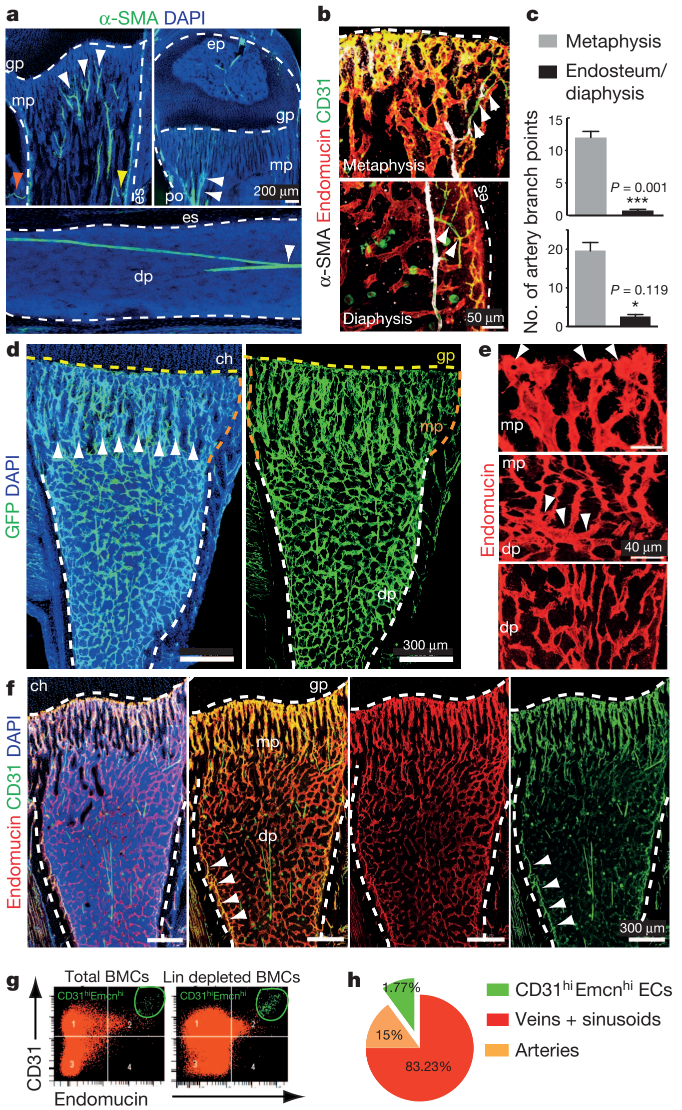


Figure 1 | Identification of bone vessel subtypes. a, Confocal images showing α-SMA$^+$ arteries (green) in thick sections (300 μm) of juvenile 4-week-old tibia. Nuclei, DAPI (4′,6-diamidino-2-phenylindole, blue). Arteries enter through the cortex (red arrowhead), terminate in endosteum (yellow), and branch in the metaphysis (mp) or diaphysis (dp) (white). Dashed lines mark growth plate (gp) or endosteum (es). **b**, α-SMA$^+$ (white) arteries and α-SMA$^-$CD31$^+$ (green) Emcn$^-$ (red) arterioles terminate in CD31$^+$Emcn$^+$ capillaries within metaphysis and endosteum (es). **c**, Quantitation of α-SMA$^+$ (top) and CD31$^+$Emcn$^-$ (bottom) arterial branch points in 4-week-old tibia. Data represent mean ± s.e.m. ($n = 6$ mice for quantification of α-SMA$^+$ arterial branch points from six independent experiments and $n = 4$ for quantification of CD31$^+$Emcn$^-$ arterial branch points; both from four independent experiments). P values, two-tailed unpaired t-tests. **d**, Representative tile scans of GFP$^+$ (green) endothelium in juvenile *Cdh5(PAC)-CreERT2*, *Rosa26-mT/mG* double heterozygous mouse tibia. Yellow dashed line marks growth plate (gp). Note distinct organization and interconnections (arrowheads) of microvessels in metaphysis (mp) and diaphysis (dp). **e**, Maximum intensity projections of endothelial cell columns in metaphysis (mp, arrowheads mark distal protrusions), which are connected (arrowheads in central panel) to highly branched sinusoids in diaphysis (dp). **f**, Confocal tile scan of juvenile 3-week-old tibia showing CD31$^+$ (green) and Emcn$^+$ (red) endothelial cells in metaphysis (mp) and endosteum (arrowheads). **g**, Representative flow cytometry dot plots showing CD31^{hi}Emcn^{hi} endothelial cells in total and lineage (Lin) depleted bone marrow cells (BMCs). **h**, Relative abundance of endothelial cell (EC) subtypes in 4-week-old long bone. CD31^{hi}Emcn^{hi} cells represent 1.77 ± 0.01% (mean ± s.d., $n = 7$ mice from three independent experiments) of total endothelial cells.

with earlier dye injection and corrosion cast experiments[10,26], endothelial tubes in the metaphysis resembled straight columns that were interconnected by distal vessel loops or arches. In contrast, diaphyseal capillaries displayed the highly branched pattern characteristic of the sinusoidal vasculature of bone marrow (Fig. 1d, e and Extended Data Fig. 4a, b). At the interface between metaphysis and diaphysis, the two vessel types were connected confirming that they were part of one continuous vascular bed (Fig. 1e). *Cdh5(PAC)-CreERT2* and *Rosa26-mT/mG* double transgenics also validated the endothelial cell specificity of Cre and absence of GFP in chondrocytes, osteoprogenitors, osteocytes or haematopoietic cells (Extended Data Fig. 4c–e).

The different vessel types were distinguishable by immunostaining with specific cell surface markers. Columnar tubes and arches in the metaphysis and endosteal endothelial cells were strongly positive for CD31 (also known as PECAM1) and Endomucin (Emcn), whereas sinusoidal vessels in the diaphysis displayed only weak CD31 staining and slightly lower Emcn expression (Fig. 1f and Extended Data Fig. 4f, g). A distinct CD31^{hi}Emcn^{hi} endothelial subset could be also identified and separated from CD31^{lo}Emcn^{lo} cells in single-cell suspensions of long bones (see Methods) (Fig. 1g). Analysis of arteries in relation to capillary subsets indicated that distal CD31$^+$ Emcn$^-$ arterioles terminated within the CD31^{hi}Emcn^{hi} endothelium (Fig. 1b). Quantitative analysis by flow cytometry showed that CD31^{hi}Emcn^{hi} cells represented only a small fraction of total endothelial cells (Fig. 1h).

These observations established the existence of spatial and phenotypic heterogeneity in the endothelium of long bone, which also applied to other skeletal elements in mice such as vertebra, sternum and calvarium (Extended Data Fig. 5a–d). Analysis of endothelial cells from other organs showed that they all, with the exception of liver, lacked a CD31^{hi}Emcn^{hi} subset (Extended Data Fig. 5e). On the basis of these findings, we propose the following terminology for bone microvessels: type H for the small CD31^{hi}Emcn^{hi} subset and type L for the CD31^{lo} Emcn^{lo} sinusoidal vessels.

Type H endothelium and osteoprogenitor cells

Immunostaining showed that Osterix$^+$ osteoprogenitors, which will give rise to osteoblasts and osteocytes[27], were selectively positioned around type H but not type L endothelium (Fig. 2a). Likewise, collagen type 1α$^+$ osteoblastic cells and Runx2$^+$ early osteoprogenitors were abundant around the CD31$^+$ columns in the metaphysis and CD31$^+$ endosteal vessels, but essentially absent from the vicinity of diaphyseal type L vessels (Fig. 2b, c). Despite the low frequency (∼1.77%) of type H endothelial cells in the bone endothelial cell fraction and ∼0.015% in total bone marrow (Fig. 1g), the majority of Runx2$^+$ (82.63 ± 1.8%), collagen 1α$^+$ (74 ± 3.3%) and Osterix$^+$ cells (70 ± 1.9%) were located directly adjacent to CD31^{hi}Emcn^{hi} vessels (Fig. 2d). Mesenchymal cells expressing platelet-derived growth factor receptor β (PDGFRβ) were also preferentially associated with type H vessels in metaphysis and endosteum (Fig. 2e).

To understand this distribution pattern of osteoblastic cells, the expression of mRNAs for secreted growth factors with known roles in osteoprogenitor survival and proliferation was analysed in freshly purified endothelial cells from long bone. *Pdgfa, Pdgfb*, *Tgfb1*, *Tgfb3* and *Fgf1* transcripts were expressed significantly higher in type H relative to type L endothelial cells (Fig. 3a). Likewise, quantitative PCR (qPCR) analysis of flushed tibiae retaining the endosteal type H endothelial cells revealed strong enrichment of the same transcripts in comparison to extracted marrow containing type L endothelial cells (Extended Data Fig. 6a, b). Accordingly, the two bone capillary endothelial cell subsets have specific expression profiles suggesting specialized functional properties.

Loss of type H endothelium during ageing

It has been previously reported that osteoblast numbers decline during ageing[28]. Due to reduced osteogenesis, bone quality and fracture healing potential deteriorate with age, whereas fracture incidences become

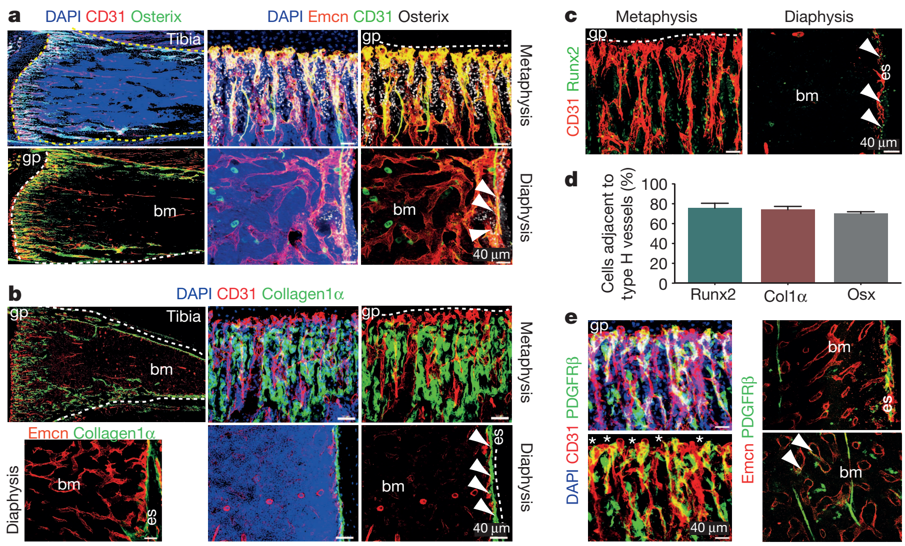


Figure 2 | Osteoprogenitor association with type H endothelial cells. **a**, Confocal images of 4-week-old tibia with indicated stainings. Growth plate (gp) and bone marrow cavity (bm) are marked. Osterix$^+$ cells are found around CD31^{hi}Emcn^{hi} vessels in metaphysis and endosteum (arrowheads). **b**, Collagen1α$^+$ cells (green) surround columnar CD31$^+$ vessels (red) but not CD31$^-$Emcn$^+$ type L bone marrow (bm) sinusoids in 3-week-old mouse tibia. **c**, Immunostained 4-week-old tibia showing association of Runx2$^+$osteoprogenitors (green) with CD31$^+$ (red) vessels in metaphysis and endosteum (es). Minimum exposure was used to capture only cells with high CD31 fluorescence in **a**–**c**. **d**, Quantitative analysis of proximity (≤ 20 µm) of Runx2$^+$, Collagen1α$^+$ (Col1α) and Osterix$^+$ (Osx) to nearest type H vessel. Mean ± s.e.m., $n = 5$ mice from three independent experiments. **e**, Maximum intensity projection of PDGFRβ$^+$ cells (green) next to CD31$^+$ (red) metaphyseal columns and distal arches (asterisks), and close to endosteal (es) type H endothelial cells. PDGFRβ also marked arteries and rare cells (arrowheads) associated with Emcn$^+$ type L bone marrow (bm) sinusoids (right).

more frequent[29]. A remarkable decline in osteoprogenitors was seen in long bone of ageing mice, accompanied by a significant decrease of transcripts for mesenchymal cell and osteoprogenitor markers and the loss of bone mass (Extended Data Fig. 6c–f). These changes correlated with pronounced reduction of type H vessels, which were much more abundant in juvenile (4-week-old) mice compared to (11-week-old) adults, and were nearly absent in aged (70-week-old) animals (Fig. 3b and Extended Data Fig. 6c). Flow cytometry analysis confirmed this age-dependent reduction of CD31^{hi}Emcn^{hi} endothelial cells, whereas the total endothelial cell number in bone did not change significantly (Fig. 3c and Extended Data Fig. 6g). Endothelial cell proliferation was high within the type H subpopulation in juvenile mice and declined rapidly in adulthood (Fig. 3d and Extended Data Fig. 6h). In contrast, the rate of type L endothelial cell proliferation did not differ significantly between juvenile and older animals (Fig. 4d). While flow cytometry had suggested the potential existence of a liver CD31^{hi}Emcn^{hi} endothelial cell subpopulation, these cells did not decline in adult and aged mice (Extended Data Fig. 6i).

Type H endothelium mediates neo-angiogenesis in bone

Genetic lineage tracing is a powerful approach for the investigation of cell amplification, differentiation and migration processes[30,31]. Single injection of 4-week-old *Cdh5(PAC)-CreERT2*, *Rosa26-mT/mG* double heterozygotes with a low dose of tamoxifen (see Methods) led to selective green fluorescent protein (GFP) expression in metaphyseal and endosteal type H endothelial cells as well as the endothelium of CD31$^+$ arteries (Fig. 4a and Extended Data Fig. 7a). The analysis of mice at day 40 after tamoxifen administration (at 9.7 weeks of age) revealed profound expansion of the GFP-positive endothelial cells, which were abundant throughout the metaphysis and diaphysis (Fig. 4a and Extended Data Fig. 7b, c). As the latter contains CD31^{lo}Emcn^{lo} vessels, this finding established that type H endothelial cells can give rise to type L endothelium. This, together with their high proliferative capacity, suggested that type H endothelial cells are hierarchically upstream of the majority of the bone sinusoidal (type L) endothelium. Given the robust labelling of *Cdh5(PAC)-CreERT2*, *Rosa26-mT/mG* double heterozygous arteries in this fate mapping approach, we cannot resolve whether type H endothelial cells give rise to arteries in postnatal bone. Arterial endothelial cells show low proliferative activity[32,33] arguing against a significant contribution of arterial GFP$^+$ cells to type L capillaries.

Exposure to irradiation is known to induce regression of sinusoidal endothelium, and a transient but substantial decline in endothelial cell number[16]. Flow cytometric quantification at 7 days post-irradiation showed a significant increase in type H endothelial cells and strong reduction of the type L subpopulation (Fig. 4b). Likewise, immunostaining of tibia sections showed the presence of CD31^{hi}Emcn^{hi} endothelial cells throughout the diaphysis (Fig. 4b). These findings are consistent with a role for CD31^{hi}Emcn^{hi} vessels in developmental and regenerative neo-angiogenesis in bone.

HIF role in angiogenesis and osteogenesis

HIF-1α controls physiological and pathological neo-angiogenesis[34]. In addition to its role in the upregulation of VEGF-A expression in hypoxic tissues, HIF-1α activity in endothelial cells is important for wound healing and tumour angiogenesis in mice[35]. To investigate HIF-1α function in the postnatal bone endothelium, inducible endothelial-specific loss-of-function mice (*Hif1a*$^{i\Delta EC}$) were generated by combining *loxP*-flanked *Hif1a* alleles (*Hif1a*$^{lox/lox}$)[35] and *Cdh5(PAC)-CreERT2* transgenics. Following tamoxifen administration from postnatal day 10 (P10) to P14, analysis of *Hif1a*$^{i\Delta EC}$ mutants at P20 revealed striking vascular defects. Whereas the amount of type H endothelium was strongly reduced in metaphysis and endosteum, the number of diaphyseal type L vessels and endothelial cells was comparable to control littermates (Fig. 4c and Extended Data Fig. 7d, e).

The von Hippel-Lindau (VHL) E3 ubiquitin protein ligase controls the stability and thereby biological activity of HIF-1α and other substrates[36]. Inducible, endothelial cell specific targeting of the murine *Vhl* gene with the same strategy as described above for *Hif1a* led to pronounced expansion of type H endothelium and metaphyseal vessel columns (Fig. 4d and Extended Data Fig. 7f). These changes were accompanied by expansion of the metaphysis, increased formation of trabecular

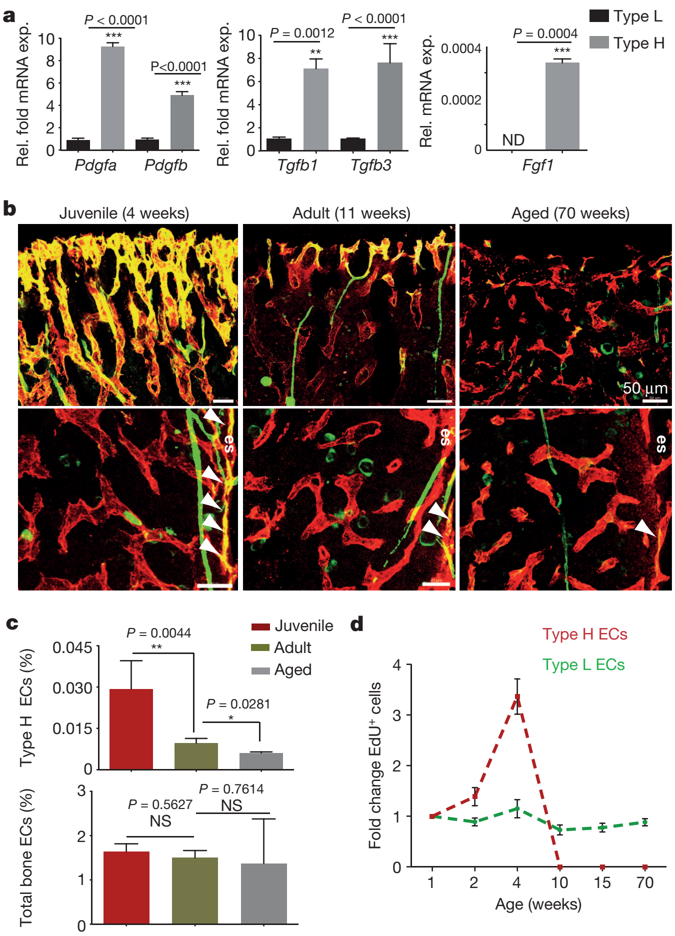


Figure 3 | Properties and age-dependent decline of type H endothelial cells. **a**, qPCR analysis of growth factor expression (normalized to *Actb*) by CD31^{hi}Emcn^{hi} endothelial cells relative to CD31^{lo}Emcn^{lo} ECs sorted from murine tibia. Data represent mean ± s.e.m. ($n = 3$ biological replicates). *P* values, two-tailed unpaired *t*-test. ND, not detectable. **b**, Confocal images of CD31 (green) and Emcn (red) immunostained tibia sections at 4 or 11 or 70 weeks. Note age-dependent decline of CD31^{hi}Emcn^{hi} endotheial cells in metaphysis (top row) and endosteum (es, arrowheads). **c**, Flow cytometric quantitation of CD31^{hi}Emcn^{hi} endothelial cells (ECs) from long bone of the indicated age group. Total endothelial cells were not significantly changed (bottom). Data represent mean ± s.e.m. ($n = 7$ mice from three independent experiments). **d**, Fold change in EdU^{+} type H and type L endothelial cell numbers in murine tibia determined by flow cytometry at indicated ages. Data represent mean ± s.e.m. ($n = 7$ mice from two independent experiments).

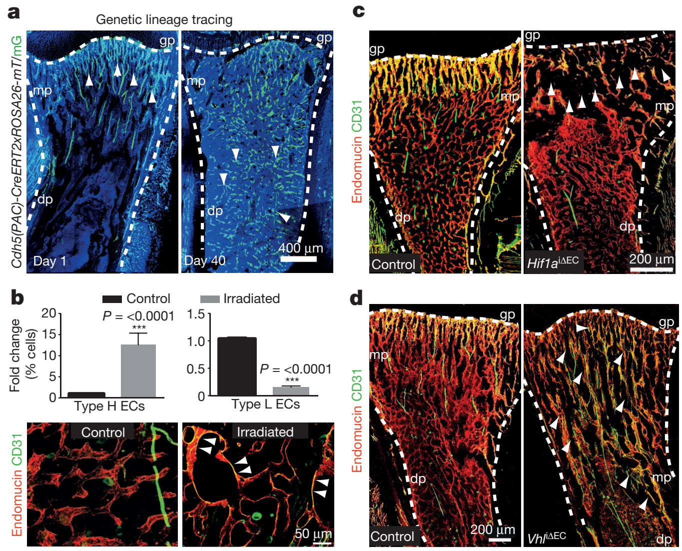


Figure 4 | Type H endothelial cells mediate bone vessel growth. **a**, Maximum intensity projection showing restricted GFP labelling (green) of arteries and CD31^{hi}Emcn^{hi} vessels in 4-week-old *Cdh5(PAC)-CreERT2, Rosa26-mT/mG* tibia at day 1 after tamoxifen administration (left). At day 40, GFP^{+} endothelial cells were found throughout the metaphysis (mp) and diaphysis (dp, arrowheads). Nuclei, DAPI (blue). **b**, Quantitation of type H and type L endothelial cells in long bone at day 7 after sublethal irradiation. Data represent mean ± s.e.m. ($n = 16$ mice, two independent experiments). *P* values, two-tailed unpaired *t*-test. Bottom panels show representative images of control and irradiated diaphysis. CD31^{hi}Emcn^{hi} endothelial cells (arrowheads) were found throughout the diaphysis after irradiation. **c**, **d**, Maximum intensity projections of immunostained, 3-week-old *Hif1a*$^{i\Delta EC}$ (**c**) or *Vhl*$^{i\Delta EC}$ tibiae (**d**) with littermate controls. Note reduction of CD31^{hi}Emcn^{hi} endothelial cells (arrowheads) in *Hif1a*$^{i\Delta EC}$ metaphysis (mp) and expansion in *Vhl*$^{i\Delta EC}$ samples. Growth plate, gp; diaphysis, dp.

bone and higher numbers of Runx2^{+} and Osterix^{+} osteoprogenitors in *Vhl*$^{i\Delta EC}$ mutants (Fig. 5a, b and Extended Data Fig. 7g). Osteoprogenitors were significantly reduced in *Hif1a*$^{i\Delta EC}$ samples (Fig. 5c, d). Likewise, blood alkaline phosphatase levels (a biomarker for osteogenesis) were elevated in *Vhl*$^{i\Delta EC}$ mutants and reduced in *Hif1a*$^{i\Delta EC}$ animals (Extended Data Fig. 8a, b). Histomorphometric analysis and microtomography (μ-CT) confirmed increased bone mass in *Vhl*$^{i\Delta EC}$ mutants without detectable change in calcitonin receptor-positive osteoclasts (Extended Data Fig. 8c–m). Parathyroid hormone and calcitonin, hormones that regulate bone remodelling, and serum calcium and phosphate levels were not significantly changed in *Hif1a*$^{i\Delta EC}$ or *Vhl*$^{i\Delta EC}$ mutants (data not shown).

To understand why endothelial-cell-specific deletion of *Hif1a* or *Vhl* primarily affected type H vessels, we analysed HIF-1α expression in 2-week-old bone, which corresponds to the time point of tamoxifen administration. Strong HIF-1α immunostaining was seen in type H but not type L endothelium (Extended Data Fig. 9a). Furthermore, analysis of freshly isolated type H cells contained significantly higher *Hif1a* transcript levels than type L endothelial cells (Extended Data Fig. 9b). Transcripts encoding HIF-1β, which heterodimerizes with HIF-1α, were not significantly different (data not shown). In endothelial cells of adult and ageing animals, *Hif1a* mRNA was slightly reduced (0.5-fold) but HIF-1α protein was no longer detectable in endothelial cells (Extended Data Fig. 9c, d), which correlated with the age-dependent loss of type H vessels. Conversely, loss of type L endothelial cells and expansion of CD31^{hi}Emcn^{hi} endothelium after irradiation (Fig. 4b) was accompanied by increased HIF-1α expression at both protein and transcript levels (Extended Data Fig. 9e, f). These findings support a key role of HIF-1α in the induction of type H endothelial cells.

Prolyl-4-hydroxylases (PHDs) modify HIF-1α and thereby mark the protein for degradation under normoxic conditions. Accordingly, PHD inhibitors, such as deferoxamine mesylate (DFM), enhance HIF-1α stability and activity[37]. Treatment of isolated primary bone endothelical cells with DFM led to upregulation of *Pecam1* and *Emcn* transcripts, but also increased the expression of growth factor messenger RNAs that were also enriched in freshly isolated CD31^{hi}Emcn^{hi} endothelial cells (Fig. 5e). Next, we tested whether DFM promotes CD31^{hi}Emcn^{hi} endothelical cells, neo-angiogenesis and osteogenesis in aged animals. Although long bones of aged, 64- to 70-week-old mice treated with vehicle control contained very few CD31^{hi}Emcn^{hi} vessels, DFM administration (see Methods) led to substantial expansion of type H endothelium (Fig. 5f and Extended Data Fig. 10a) and emergence of vessel-associated Osterix^{+} cells (Fig. 5g and Extended Data Fig. 10b–d). Expression levels of osteoprogenitors and osteoblasts markers were significantly increased in long bone and calvarium of DFM-treated animals (Fig. 5h and Extended Data Fig. 10e). Furthermore, μ-CT examination showed that 6 weeks of DFM treatment led to significantly increased bone mass (Fig. 5i, j). Although the activity of DFM is not restricted to endothelial cells and is likely to affect multiple cell populations, the findings above argue

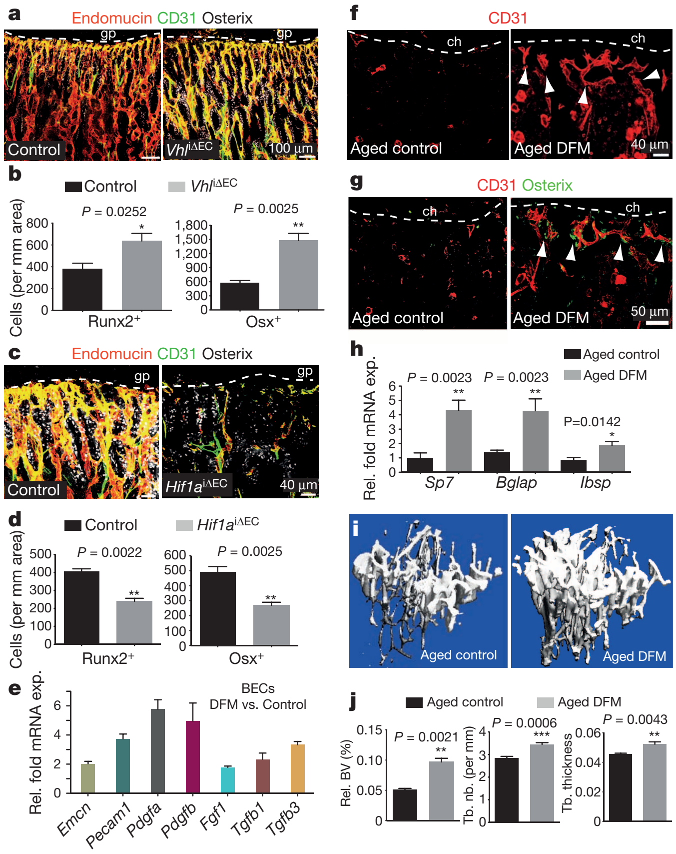


Figure 5 | Type H endothelial cells couple angiogenesis and osteogenesis. **a**, Confocal images of CD31 (green), Endomucin (red) and Osterix (white) immunostained, 3-week-old $Vhl^{i\Delta EC}$ and control tibiae. **b**, Quantitation of Runx2$^+$ and Osterix$^+$ (Osx) cells in $Vhl^{i\Delta EC}$ mutants and controls. Data represent mean ± s.e.m. ($n = 5$ mice in two independent experiments). P values, two-tailed unpaired t-test. **c**, Maximum intensity projections of 3-week-old $Hif1a^{i\Delta EC}$ and control tibia stained for CD31 (green), Endomucin (red) and Osterix (white). Growth plate, gp. **d**, Quantitation of Runx2$^+$ and Osterix$^+$ cells in $Hif1a^{i\Delta EC}$ mutant and control long bone. Data represent mean ± s.e.m. ($n = 5$ mice in three independent experiments). P values, two-tailed unpaired t-test. **e**, qPCR mRNA analysis of DFM or vehicle-treated bone endothelial cells (BECs). Data represent mean ± s.e.m. ($n = 3$ biological replicates). Note induction of growth factor expression by DFM. **f**, **g**, Representative images of CD31 (red, **f**, **g**) and Osterix (green, **g**) stained tibia sections from aged DFM-treated and control mice. Low intensity projection shows only CD31^{hi} cells. DFM induces CD31^{hi} vessels and Osterix$^+$ osteoprogenitors. Chondrocytes, ch. **h**, qPCR analysis of *Sp7* (encoding Osterix), *Bglap* (bone gamma-carboxyglutamate protein) and *Ibsp* (integrin-binding sialoprotein) mRNA expression relative to *Actb* in the aged DFM-treated or vehicle control femur. Data represent mean ± s.e.m. ($n = 6$–8 mice in four independent experiments). P values, two-tailed unpaired t-test. **i**, Representative µ-CT images of aged DFM-treated and control tibiae. **j**, Quantitative µ-CT analysis of relative bone volume (Rel. BV; bone volume/total volume), trabecular number (Tb. nb.), and trabecular thickness (Tb. thickness) in proximal tibia. Data represent mean ± s.e.m. ($n = 5$ mice in two independent experiments). P values, two-tailed unpaired t-test.

for crucial roles of endothelial HIF in controlling bone angiogenesis, type H vessel abundance, endothelial growth factor expression and osteogenesis.

Discussion

The mammalian skeleton sustains stem and progenitor cell populations allowing lifelong haematopoiesis and osteogenesis[38,39]. These processes rely on niche microenvironments for which a number of critical cell types and molecular factors have been identified[40,41]. In the case of bone formation, mesenchymal stem cells generate new osteoblasts and their progenitors, which are essential for bone homeostasis and fracture healing[42,43]. Blood vessel growth and the invasion of osteoprogenitors appear coupled in bone fractures[8] but the communication between endothelial and osteoblastic cells remains little understood. Likewise, the exact architecture of the bone vasculature or the processes mediating its growth had remained unknown. Some of these key questions are now resolved in this manuscript. The finding that capillaries in the skeletal system of mice can be subdivided into type H and type L endothelium on the basis of morphological, molecular and functional criteria should be very useful for future studies in basic and medical research. CD31^{hi}Emcn^{hi} capillaries at the distal end of the arterial network in bone might represent the central building block of a metabolically specialized tissue environment with privileged access to oxygen and nutrients, which probably influences the growth potential and metabolism of other cell types. This is not only relevant for osteoblastic cells but potentially also for haematopoietic stem and progenitor cells, which preferentially home to the metaphysis after transplantation[44]. However, there is also evidence that the bulk of haematopoietic stem cells are exposed to low oxygen, whereas committed and differentiating haematopoietic progenitors reside in relatively well-oxygenated environments[45]. The latter is consistent with our recent description of the hemosphere, a specialized microenvironment found in the mouse metaphysis that enables the clonal expansion of haematopoietic cells[46].

We also propose that type H endothelial cells mediate local growth of the vasculature and provide niche signals for perivascular osteoprogenitors. Type H vessel formation and the expression of potential angiocrine factors for osteoblastic cells are enhanced by HIF and, as we show in the accompanying manuscript[47], by Notch signalling. Thus, the abundance of CD31^{hi}Emcn^{hi} endothelial cells may be useful as diagnostic read-out for the growth status of the bone vasculature and its pro-osteogenic capacity. Our results also indicate that specific molecular pathways can be used to boost type H vessel formation and osteogenesis (Extended Data Fig. 1b). This might be of great importance for conditions involving compromised fracture healing or loss of bone mass. Ageing and post-menopausal oestrogen deficiency are major risk factors for osteoporosis[48], and oestrogen can promote angiogenesis[49]. Accordingly, decline of type H vessels and the concomitant reduction of osteoprogenitor cells could potentially offer a compelling explanation for the loss of bone mass during ageing and might enable therapeutic improvement of osteogenesis in elderly people.

METHODS SUMMARY

Genetically modified and aged mice. C57BL/6J males were used for experiments at the age of 2–4 weeks (juvenile), 8–12 weeks (adult) and 57–70 weeks (aged). Endothelial-cell-specific gene deletions were generated using *Cdh5(PAC)-CreERT2* transgenic mice, which were interbred with conditional mutants carrying *loxP*-flanked *Hif1a* ($Hif1a^{lox/lox}$)[35]. To induce Cre activity and gene inactivation, pups were injected with 500 µg tamoxifen (Sigma, T5648) intraperitoneally (i.p.) every-day from P10 to P14. Femurs and tibiae from $Cdh5(PAC)\text{-}CreERT2^{T/+}$ $Hif1a^{lox/lox}$ ($Hif1a^{i\Delta EC}$) mutants and Cre-negative ($Hif1a^{lox/lox}$) controls were collected on P20 or P37 after euthanasia. The same approach was used for experiments involving conditional *Vhl* mutant mice[50]. For genetic labelling of metaphyseal vasculature, $Cdh5(PAC)\text{-}CreERT2^{T/+}$ mice were mated with R26-mG/mT reporters. At P29, $Cdh5(PAC)\text{-}CreERT2^{T/+}/R26\text{-}mG/mT^{T/+}$ mice received a single dose of 50 µg tamoxifen and were analysed either 1 or 40 days later.

All animals were genotyped by PCR. Protocols and primer sequences are provided upon request. When indicated, adult wild-type mice were whole-body irradiated with a single dose of 900 rads (Gammacell irradiator) and euthanized 7 days later.

For DFM treatment, freshly prepared deferoxamine mesylate (Sigma) in water (15 mg ml^{-1} per mouse) was injected i.p. every other day for 4 weeks or, for µ-CT analysis, 5 weeks. Control animals received the same amount of sterile water.

For labelling of proliferating cells, mice were intraperitoneally injected with 1.6 mg per kg weight of EdU (Invitrogen) 2 h before euthanasia. Tibiae were immediately collected and processed. Bone marrow cells and bone sections were stained for EDU using Click-iT chemistry following the manufacturer's instructions (Invitrogen).

Experiments involving animals were performed according to the institutional guidelines and laws, following protocols approved by local animal ethics committees.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 July 2013; accepted 11 February 2014.

Published online 12 March 2014.

1. Tashiro, Y. *et al.* Inhibition of PAI-1 induces neutrophil-driven neoangiogenesis and promotes tissue regeneration via production of angiocrine factors in mice. *Blood* **119,** 6382–6393 (2012).
2. Red-Horse, K., Crawford, Y., Shojaei, F. & Ferrara, N. Endothelium-microenvironment interactions in the developing embryo and in the adult. *Dev. Cell* **12,** 181–194 (2007).
3. Ding, L., Saunders, T. L., Enikolopov, G. & Morrison, S. J. Endothelial and perivascular cells maintain haematopoietic stem cells. *Nature* **481,** 457–462 (2012).
4. Butler, J. M., Kobayashi, H. & Rafii, S. Instructive role of the vascular niche in promoting tumour growth and tissue repair by angiocrine factors. *Nature Rev. Cancer* **10,** 138–146 (2010).
5. Ribatti, D., Nico, B., Vacca, A., Roncali, L. & Dammacco, F. Endothelial cell heterogeneity and organ specificity. *J. Hematother. Stem Cell Res.* **11,** 81–90 (2002).
6. Garlanda, C. & Dejana, E. Heterogeneity of endothelial cells. Specific markers. *Arterioscler. Thromb. Vasc. Biol.* **17,** 1193–1202 (1997).
7. Eshkar-Oren, I. *et al.* The forming limb skeleton serves as a signaling center for limb vasculature patterning via regulation of *Vegf*. *Development* **136,** 1263–1272 (2009).
8. Maes, C. *et al.* Increased skeletal VEGF enhances β-catenin activity and results in excessively ossified bones. *EMBO J.* **29,** 424–441 (2010).
9. Trueta, J. & Buhr, A. J. The vascular contribution to osteogenesis. V. the vasculature supplying the epiphysial cartilage in rachitic rats. *J. Bone Joint Surg. Br.* **45,** 572–581 (1963).
10. Trueta, J. & Morgan, J. D. The vascular contribution to osteogenesis. I. Studies by the injection method. *J. Bone Joint Surg. Br.* **42-B,** 97–109 (1960).
11. Glowacki, J. Angiogenesis in fracture repair. *Clin. Orthop. Relat. Res.* **355,** (suppl.), S82–S89 (1998).
12. Maes, C. *et al.* Osteoblast precursors, but not mature osteoblasts, move into developing and fractured bones along with invading blood vessels. *Dev. Cell* **19,** 329–344 (2010).
13. Burkhardt, R. *et al.* Changes in trabecular bone, hematopoiesis and bone marrow vessels in aplastic anemia, primary osteoporosis, and old age: a comparative histomorphometric study. *Bone* **8,** 157–164 (1987).
14. Lu, C. *et al.* Effect of age on vascularization during fracture repair. *J. Orthop. Res.* **26,** 1384–1389 (2008).
15. Kataoka, M. & Tavassoli, M. Identification of lectin-like substances recognizing galactosyl residues of glycoconjugates on the plasma membrane of marrow sinus endothelium. *Blood* **65,** 1163–1171 (1985).
16. Kopp, H. G., Avecilla, S. T., Hooper, A. T. & Rafii, S. The bone marrow vascular niche: home of HSC differentiation and mobilization. *Physiology (Bethesda)* **20,** 349–356 (2005).
17. Hooper, A. T. *et al.* Engraftment and reconstitution of hematopoiesis is dependent on VEGFR2-mediated regeneration of sinusoidal endothelial cells. *Cell Stem Cell* **4,** 263–274 (2009).
18. Nombela-Arrieta, C. *et al.* Quantitative imaging of haematopoietic stem and progenitor cell localization and hypoxic status in the bone marrow microenvironment. *Nature Cell Biol.* **15,** 533–543 (2013).
19. Trueta, J. & Harrison, M. H. The normal vascular anatomy of the femoral head in adult man. *J. Bone Joint Surg. Br.* **35-B,** 442–461 (1953).
20. Crock, H. V. A revision of the anatomy of the arteries supplying the upper end of the human femur. *J. Anat.* **99,** 77–88 (1965).
21. Ullah, M. S., Davies, A. J. & Halestrap, A. P. The plasma membrane lactate transporter MCT4, but not MCT1, is up-regulated by hypoxia through a HIF-1α-dependent mechanism. *J. Biol. Chem.* **281,** 9030–9037 (2006).
22. Boado, R. J. & Pardridge, W. M. Glucose deprivation and hypoxia increase the expression of the GLUT1 glucose transporter via a specific mRNA cis-acting regulatory element. *J. Neurochem.* **80,** 552–554 (2002).
23. Rodríguez, J. *et al.* ERK1/2 MAP kinases promote cell cycle entry by rapid, kinase-independent disruption of retinoblastoma-lamin A complexes. *J. Cell Biol.* **191,** 967–979 (2010).
24. Wang, Y. *et al.* Ephrin-B2 controls VEGF-induced angiogenesis and lymphangiogenesis. *Nature* **465,** 483–486 (2010).
25. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45,** 593–605 (2007).
26. Skawina, A., Litwin, J. A., Gorczyca, J. & Miodonski, A. J. The vascular system of human fetal long bones: a scanning electron microscope study of corrosion casts. *J. Anat.* **185,** 369–376 (1994).
27. Nakashima, K. *et al.* The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. *Cell* **108,** 17–29 (2002).
28. Lips, P., Courpron, P. & Meunier, P. J. Mean wall thickness of trabecular bone packets in the human iliac crest: changes with age. *Calcif. Tissue Res.* **26,** 13–17 (1978).
29. Smith, D. M., Khairi, M. R. & Johnston, C. C., Jr. The loss of bone mineral with aging and its relationship to risk of fracture. *J. Clin. Invest.* **56,** 311–318 (1975).
30. Šale, S., Lafkas, D. & Artavanis-Tsakonas, S. Notch2 genetic fate mapping reveals two previously unrecognized mammary epithelial lineages. *Nature Cell Biol.* **15,** 451–460 (2013).
31. Zovein, A. C. *et al.* Fate tracing reveals the endothelial origin of hematopoietic stem cells. *Cell Stem Cell* **3,** 625–636 (2008).
32. Swift, M. R. & Weinstein, B. M. Arterial-venous specification during development. *Circ. Res.* **104,** 576–588 (2009).
33. Helisch, A. & Schaper, W. Arteriogenesis: the development and growth of collateral arteries. *Microcirculation* **10,** 83–97 (2003).
34. Pugh, C. W. & Ratcliffe, P. J. Regulation of angiogenesis by hypoxia: role of the HIF system. *Nature Med.* **9,** 677–684 (2003).
35. Tang, N. *et al.* Loss of HIF-1α in endothelial cells disrupts a hypoxia-driven VEGF autocrine loop necessary for tumorigenesis. *Cancer Cell* **6,** 485–495 (2004).
36. Tanimoto, K., Makino, Y., Pereira, T. & Poellinger, L. Mechanism of regulation of the hypoxia-inducible factor-1α by the von Hippel-Lindau tumor suppressor protein. *EMBO J.* **19,** 4298–4309 (2000).
37. Jones, D. T. & Harris, A. L. Identification of novel small-molecule inhibitors of hypoxia-inducible factor-1 transactivation and DNA binding. *Mol. Cancer Ther.* **5,** 2193–2202 (2006).
38. Kuznetsov, S. A. *et al.* The interplay of osteogenesis and hematopoiesis: expression of a constitutively active PTH/PTHrP receptor in osteogenic cells perturbs the establishment of hematopoiesis in bone and of skeletal stem cells in the bone marrow. *J. Cell Biol.* **167,** 1113–1122 (2004).
39. Bianco, P. Bone and the hematopoietic niche: a tale of two stem cells. *Blood* **117,** 5281–5288 (2011).
40. Long, M. W. Osteogenesis and bone-marrow-derived cells. *Blood Cells Mol. Dis.* **27,** 677–690 (2001).
41. Yin, T. & Li, L. The stem cell niches in bone. *J. Clin. Invest.* **116,** 1195–1201 (2006).
42. Shapiro, F. Bone development and its relation to fracture repair. The role of mesenchymal osteoblasts and surface osteoblasts. *Eur. Cell. Mater.* **15,** 53–76 (2008).
43. Park, D. *et al.* Endogenous bone marrow MSCs are dynamic, fate-restricted participants in bone maintenance and regeneration. *Cell Stem Cell* **10,** 259–272 (2012).
44. Ellis, S. L. *et al.* The relationship between bone, hemopoietic stem cells, and vasculature. *Blood* **118,** 1516–1524 (2011).
45. Parmar, K., Mauch, P., Vergilio, J. A., Sackstein, R. & Down, J. D. Distribution of hematopoietic stem cells in the bone marrow according to regional hypoxia. *Proc. Natl Acad. Sci. USA* **104,** 5431–5436 (2007).
46. Wang, L. *et al.* Identification of a clonally expanding haematopoietic compartment in bone marrow. *EMBO J.* **32,** 219–230 (2013).
47. Ramasamy, S. K., Kusumbe, A. P. & Adams, R. H. Endothelial Notch activity promotes angiogenesis and osteogenesis in bone. *Nature* http://dx.doi.org/10.1038/nature13146 (this issue).
48. Heaney, R. P. Pathophysiology of osteoporosis. *Endocrinol. Metab. Clin. North Am.* **27,** 255–265 (1998).
49. Losordo, D. W. & Isner, J. M. Estrogen and angiogenesis: A review. *Arterioscler. Thromb. Vasc. Biol.* **21,** 6–12 (2001).
50. Haase, V. H., Glickman, J. N., Socolovsky, M. & Jaenisch, R. Vascular tumors in livers with targeted inactivation of the von Hippel-Lindau tumor suppressor. *Proc. Natl Acad. Sci. USA* **98,** 1583–1588 (2001).

Acknowledgements We thank A. Medvinsky for kindly providing *Flk1-GFP* mice, M. Stehling for endothelial cell sorting and A. Borgscheiper for technical assistance. Funding was provided by the Max Planck Society, the University of Münster, the DFG cluster of excellence 'Cells in Motion' and the European Research Council (AdG 339409 AngioBone).

Author Contributions A.P.K., S.K.R. and R.H.A. designed experiments and interpreted results. A.P.K and S.K.R performed all experiments. A.P.K and R.H.A. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.H.A. (ralf.adams@mpi-muenster.mpg.de).

METHODS

Genetically modified and aged mice. C57BL/6J males were used for experiments at the age of 2–4 weeks (juvenile), 8–12 weeks (adult) and 57–70 weeks (aged). Endothelial-cell-specific gene deletions were generated using *Cdh5(PAC)-CreERT2* transgenic mice, which were interbred with conditional mutants carrying *loxP*-flanked *Hif1a* (*Hif1a*$^{lox/lox}$)[35]. To induce Cre activity and gene inactivation, pups were injected with 500 μg tamoxifen (Sigma, T5648) intraperitoneally (i.p.) every-day from P10 to P14. Femurs and tibiae from *Cdh5(PAC)-CreERT2*$^{T/+}$ *Hif1a*$^{lox/lox}$ (*Hif1a*$^{i\Delta EC}$) mutants and Cre-negative (*Hif1a*$^{lox/lox}$) controls were collected on P20 or P37 after euthanasia. The same approach was used for experiments involving conditional *Vhl* mutant mice[50]. For genetic labelling of metaphyseal vasculature, *Cdh5(PAC)-CreERT2*$^{T/+}$ mice were mated with R26-mG/mT reporters. At P29, double heterozygous *Cdh5(PAC)-CreERT2*$^{T/+}$/*R26-mG/mT*$^{T/+}$ offspring received a single dose of 50 μg tamoxifen and was analysed either 1 or 40 days later.

All animals were genotyped by PCR. Protocols and primer sequences are provided upon request. When indicated, adult wild-type mice were whole-body irradiated with a single dose of 900 rads (Gammacell irradiator) and euthanized 7 days later.

For DFM treatment, freshly prepared deferoxamine mesylate (Sigma) in water (15 mg ml^{-1} per mouse) was injected i.p. every other day for 4 weeks or, for μ-CT analysis, 5 weeks. Control animals received the same amount of sterile water.

For labelling of proliferating cells, mice were intraperitoneally injected with 1.6 mg per kg weight of EdU (Invitrogen) 2 h before euthanasia. Tibiae were immediately collected and processed. Bone marrow cells and bone sections were stained for EDU using Click-iT chemistry following the manufacturer's instructions (Invitrogen).

Experiments involving animals were performed according to the institutional guidelines and laws, following protocols approved by local animal ethics committees.

Bone immunohistochemistry. Freshly dissected bone tissues collected from wild-type mice or from mutants and their control littermates were immediately fixed in ice-cold 4% paraformaldehyde solution for 4 h. Decalcification was carried out with 0.5 M EDTA at 4 °C with constant shaking and decalcified bones were immersed into 20% sucrose and 2% polyvinylpyrrolidone (PVP) solution for 24 h. Finally, the tissues were embedded and frozen in 8% gelatin (porcine) in presence of 20% sucrose and 2% PVP. For immunofluorescent stainings and morphological analyses, sections were generated using low-profile blades on a Leica CM3050 cryostat.

For phenotypic analysis, mutant and littermate control samples were always processed, sectioned, stained, imaged and analysed together at the same conditions and settings. For immunostaining, bone sections were air-dried, permeabilised for 10 min in 0.3% Triton X-100, blocked in 5% donkey serum at room temperature for 30 min, and probed with the primary antibodies diluted in 5% donkey serum in PBS for 2 h at room temperature (RT) or overnight at 4 °C.

The following primary antibodies were used: Endomucin (sc-65495, Santa Cruz, diluted 1:100), Pecam1 conjugated to Alexa Fluor 488 (FAB3628G, R&D Systems, 1:100), Pecam1 (553370, BD Pharmingen, 1:100), Osterix (sc-22536-R, Santa Cruz, 1:200), Runx2 (MAB2006, R&D Systems, 1:200), α-SMA-Cy3 (C6198, Sigma, 1:100), Hif1-α (ab65979, Abcam, 1:100), Collagen type 1 (AB675P; Millipore, 1:200), Osteopontin (AF808, R&D Systems, 1:200), Mct4 (sc-50329, Santa Cruz, diluted 1:100), Glut1 (07-1401, Millipore, 1:100), phospho-p44/42 MAPK (4370, Cell Signaling, 1:75), Calcitonin receptor (ab11042, Abcam, 1:75), Biotin-conjugated CD45 (553077, Becton Dickinson, 1:100), and Hif1-α (ab65979, Abcam, 1:100). After primary antibody incubation, sections were washed with PBS for three times and incubated with appropriate Alexa Fluor-coupled secondary antibodies (1:400, Molecular Probes) for 1 h at room temperature. Nuclei were counterstained with DAPI. Sections were thoroughly washed with PBS before mounting them using FluoroMount-G (Southern Biotech). Finally, coverslips were sealed with nail polish.

For metabolic labelling with the hypoxia probe pimonidazole (Pimo, Hypoxyprobe Inc.), mutant and control mice were intraperitoneally injected with 60 mg per kg Pimo for 2 h before euthanasia. Metabolized Pimo was detected by a rabbit antiserum against the non-oxidized, protein-conjugated form of pimonidazole (Hypoxyprobe Inc.).

Image acquisition and quantitative analysis. Immunofluorescent stainings were analysed at high resolution with a Zeiss laser scanning confocal microscope, LSM-780. *Z*-stacks of images were processed and 3D-reconstructed with Imaris software (version 7.00, Bitplane). Imaris, Photoshop and Illustrator (Adobe) software was used for image processing in compliance with *Nature*'s guide for digital images. All quantifications were done with ImageJ, Imaris and Volocity software on high-resolution confocal images.

Statistical analysis. All data are presented as mean ± s.e.m. The significance of difference in the mean values was determined using two-tailed Student's *t* test unless indicated otherwise. $P < 0.05$ was considered significant. All statistical analysis were performed using GraphPad Prism software. No randomization or blinding was used and no animals were excluded from analysis. Sample sizes were selected on the basis of previous experiments. Several independent experiments were performed to guarantee reproducibility of findings.

Culture of endothelial cells from bone. Tibiae and femurs from wild-type mice or mice expressing endothelial specific reporters (that is, *Flk1-GFP*[51] and *Cdh5(PAC)-CreERT2*$^{T/+}$/R26-mG/mT$^{d/+}$) were collected in sterile Ca^{2+} and Mg^{2+} free PBS, crushed with mortar and pestle, and digested with collagenase A (Sigma) to obtain a single-cell suspension. Endothelial cells were then MACS sorted using Endomucin antibody (cat. no. SC-65495) and Dynabeads sheep anti-Rat IgG (Invitrogen). Sorted endothelial cells were then plated on dishes coated with fibronectin and cultured in endothelial cell growth medium (EBM-2, Clonetics; Lonza) supplemented with EGM-2 SingleQuots (CC-4176, Clonetic; Lonza). At first passage, cells were again MACS sorted with Endomucin antibody and plated for culture. Cells were fed every third day and passaged upon confluency. Cultures were maintained at 37 °C with 5% CO_2 in a humidified atmosphere. For DFM treatment and subsequent qPCR analysis, endothelial cell cultures between passage 2 and 5 were used. Cells were treated with DFM (6.25 mg ml^{-1} of culture medium) for the duration of 7 h and subsequently lysed in lysis buffer of RNeasy Mini Kit (Qiagen) for qPCR analysis.

Quantitative RT–PCR. For the analysis of mRNA expression levels in type H or type L endothelium, CD31^{hi}Emcn^{hi} and CD31^{lo}Emcn^{lo} cells were sorted by FACS directly into the lysis buffer of the RNeasy Mini Kit (Qiagen). Total RNA was isolated according to manufacturer's protocol. A total of 100 ng RNA per reaction was used to generate cDNA with the iScript cDNA Synthesis System (Bio-Rad). Quantitative PCR (qPCR) was performed using TaqMan gene expression assays on ABI PRISM 7900HT Sequence Detection System. The FAM-conjugated TaqMan probes *Tgfb1*, *Tgfb3*, *Fgf1*, *Pdgfa*, and *Pdgfb* were used along with TaqMan Gene Expression Master Mix (Applied Biosystems). Gene expression assays were normalized to endogenous VIC-conjugated *Actb* probes as standard. For analysis of mRNA expression levels from whole bones, dissected femurs were immediately crushed finely, digested with collagenase, centrifuged to obtain a pellet, which was then lysed in lysis buffer of RNeasy Mini Kit (Qiagen). A total of 500 ng RNA per reaction was used to generate cDNA with the iScript cDNA Synthesis System (Bio-Rad), which was further processed as described above. FAM-conjugated TaqMan probes *Ibsp*, *Bglap*, *Sp7*, *Cspg4*, *Pdgfrb*, *Runx2* and *Sp7* were used along with TaqMan Gene Expression Master Mix (Applied Biosystems) to perform qPCR. For the mRNA expression analysis of flushed bone and bone collar/endosteum, FAM-conjugated TaqMan probes *Tgfb1*, *Tgfb3*, *Fgf1*, *Pdgfa* and *Pdgfb* were used. Gene expression was normalized to the endogenous VIC-conjugated *Actb* probes.

Flow cytometry. For flow cytometric analysis and sorting of type H and type L endothelial cells, tibiae and femurs were collected and cleaned thoroughly to remove the adherent muscles. The epiphysis was removed and only the metaphysis and diaphysis regions were processed. Tibias were then crushed in ice-cold PBS with mortar and pestle. Whole bone marrow was digested with collagenase incubation at 37 °C for 20 min. Equal number of cells were then subjected to immunostaining with endomucin antibody (Santa Cruz, sc-65495) for 45 min. After washing, cells were stained with APC-conjugated CD31 antibody (R&D Systems, FAB3628A) for 45 min. After washing, cells were acquired on BD FACS Canto flow cytometer and analysed using BD FACSDiva software (Version 6.0, BD Bioscience). Cell sorting was performed with FACS Aria II.

For demarcating and sorting CD31^{hi}Emcn^{hi} endothelial cells, first standard quadrant gates were set, subsequently to differentiate CD31^{hi}Emcn^{hi} cells from the total double positive cells in quadrant 2 gates were arbitrarily set at $>10^4$ log Fl-4 (CD31-APC) fluorescence and $>10^4$ log Fl-2 (Endomucin-PE) fluorescence.

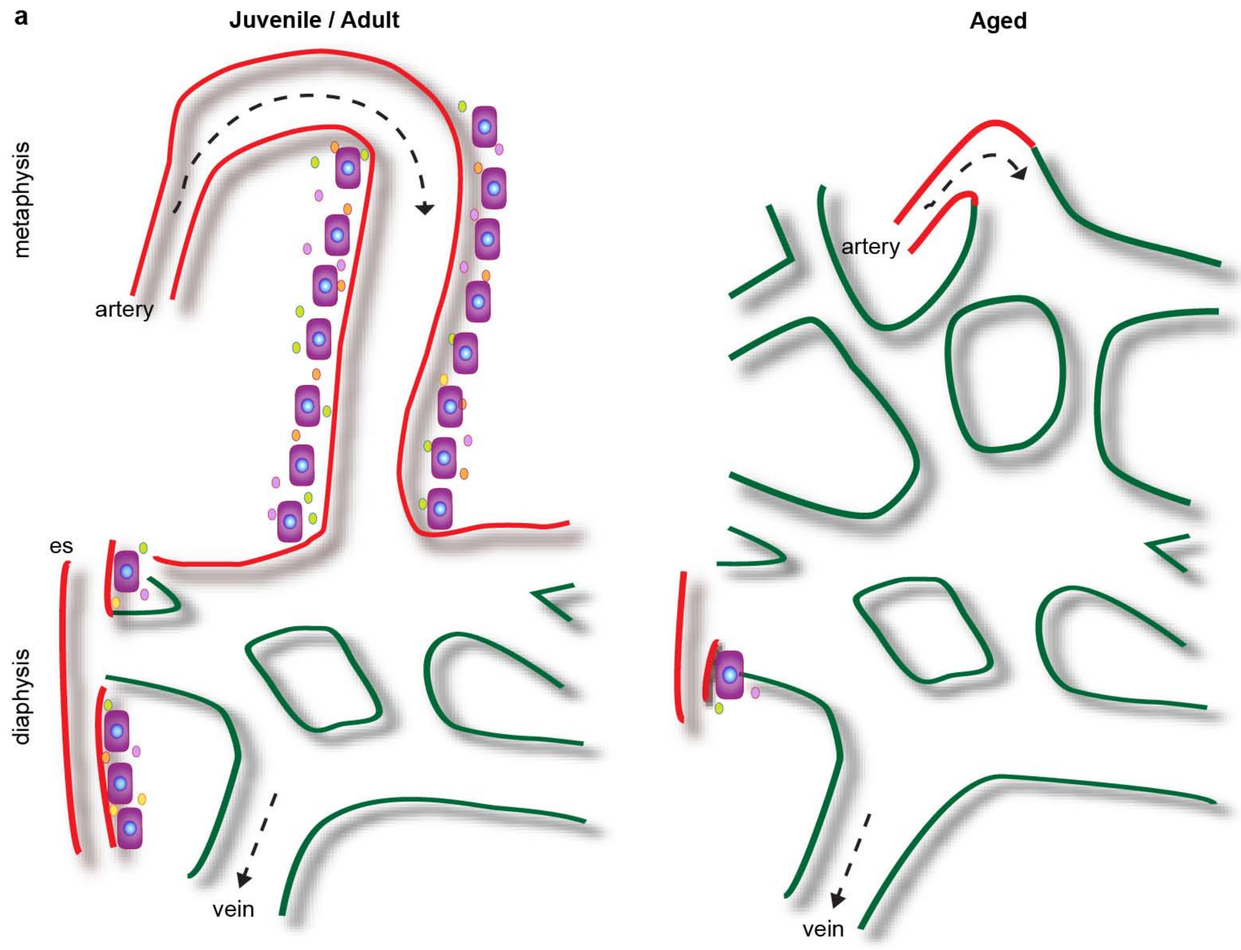
For the analysis of total ECs in bone, tibiae were processed as described above to obtain single cell suspensions, which were stained with biotin-coupled CD45 (BD, 553077) or Ter119 (BD, 559971) antibodies for 45 min. After washing in PBS, cells stained with Streptavidin PE-Cy5 (BD, 554062) and Alexa Fluor488-conjugated CD31 (R&D Systems, FAB3628G) antibodies for 45 min. After washing, cells were acquired on a FACS Canto flow cytometer and analysed using FACSDiva software (Version 6.0, BD Bioscience). Total bone ECs were quantified as CD31^{+} CD45^{-}Ter119^{-}. Endomucin was used to distinguish Emcn-negative arterial endothelial cells from Emcn^{+} sinusoidal and venous cells.

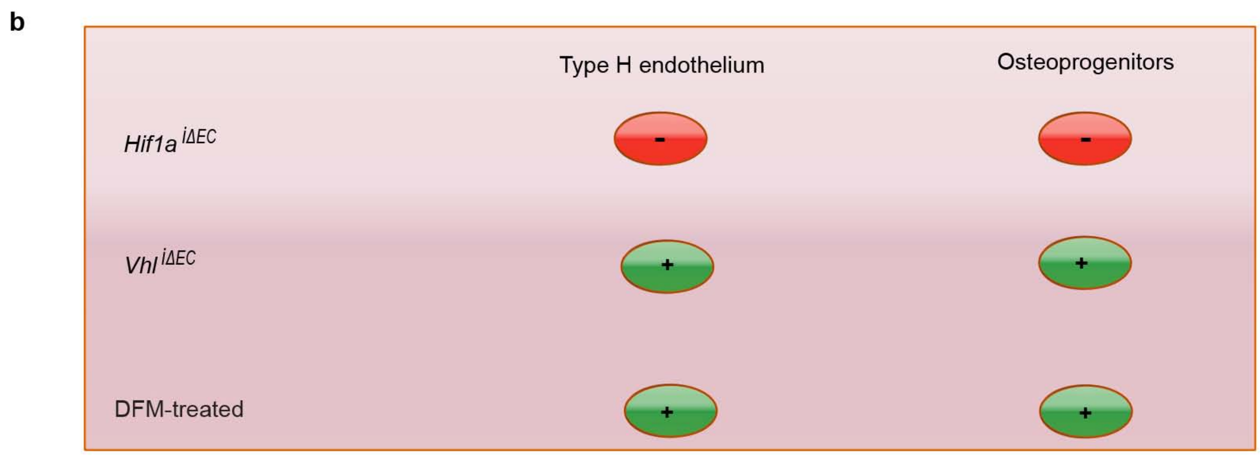
Micro-CT analysis and histomorphometry. Tibiae were collected, and the attached soft tissue was removed thoroughly and fixed in 4% paraformaldehyde. The fixed tibiae were analysed using micro-CT (μCT 35) and software IPL V5.15 at Scanco Medical AG, Switzerland. A voxel size of 12 μm was chosen in all three spatial dimensions. For each sample, 148 from 232 slices were evaluated covering a total of 1.776 mm at a voltage of 70 kVp, intensity 114 μA, integration time 1200 ms.

For calcein double labelling, mice were given intraperitoneal injections of 10mg per kg calcein (Sigma, C0875) dissolved in 2% sodium bicarbonate solution at day 10 and day 3 before euthanasia. Undecalcified bones were fixed in 4% PFA, embedded in 8% gelatin and 2% PVP and cryosectioned. Sections were stained with von Kossa method to assess mineralized bone

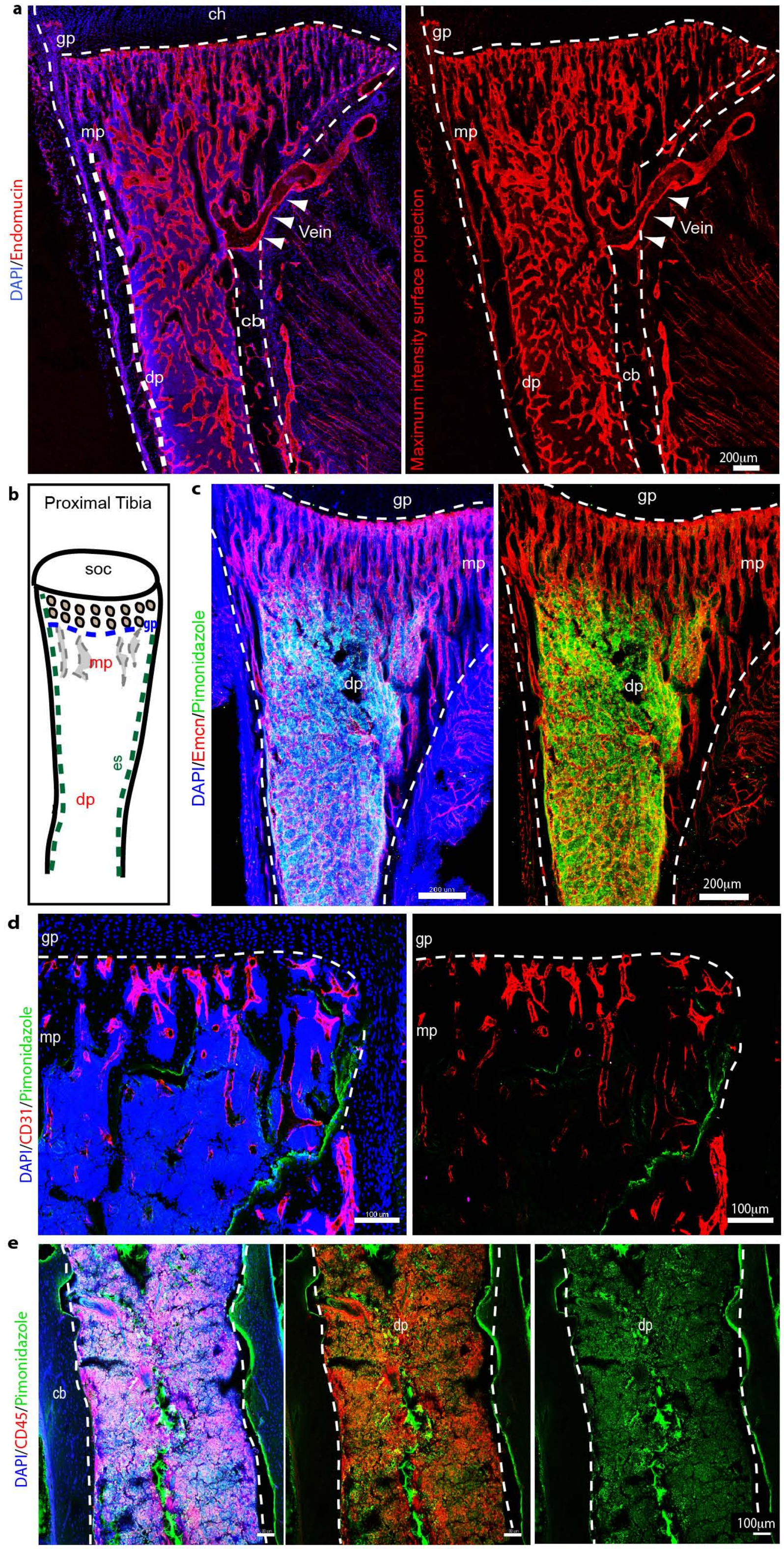
Osteoclast surface/bone surface (Oc. S/BS; %) and osteoclast number/bone perimeter (No. Oc./B. Pm) were calculated based on Calcitonin receptor staining of bone sections.

51. Xu, Y. *et al.* Neuropilin-2 mediates VEGF-C-induced lymphatic sprouting together with VEGFR3. *J. Cell Biol.* **188,** 115–130 (2010).

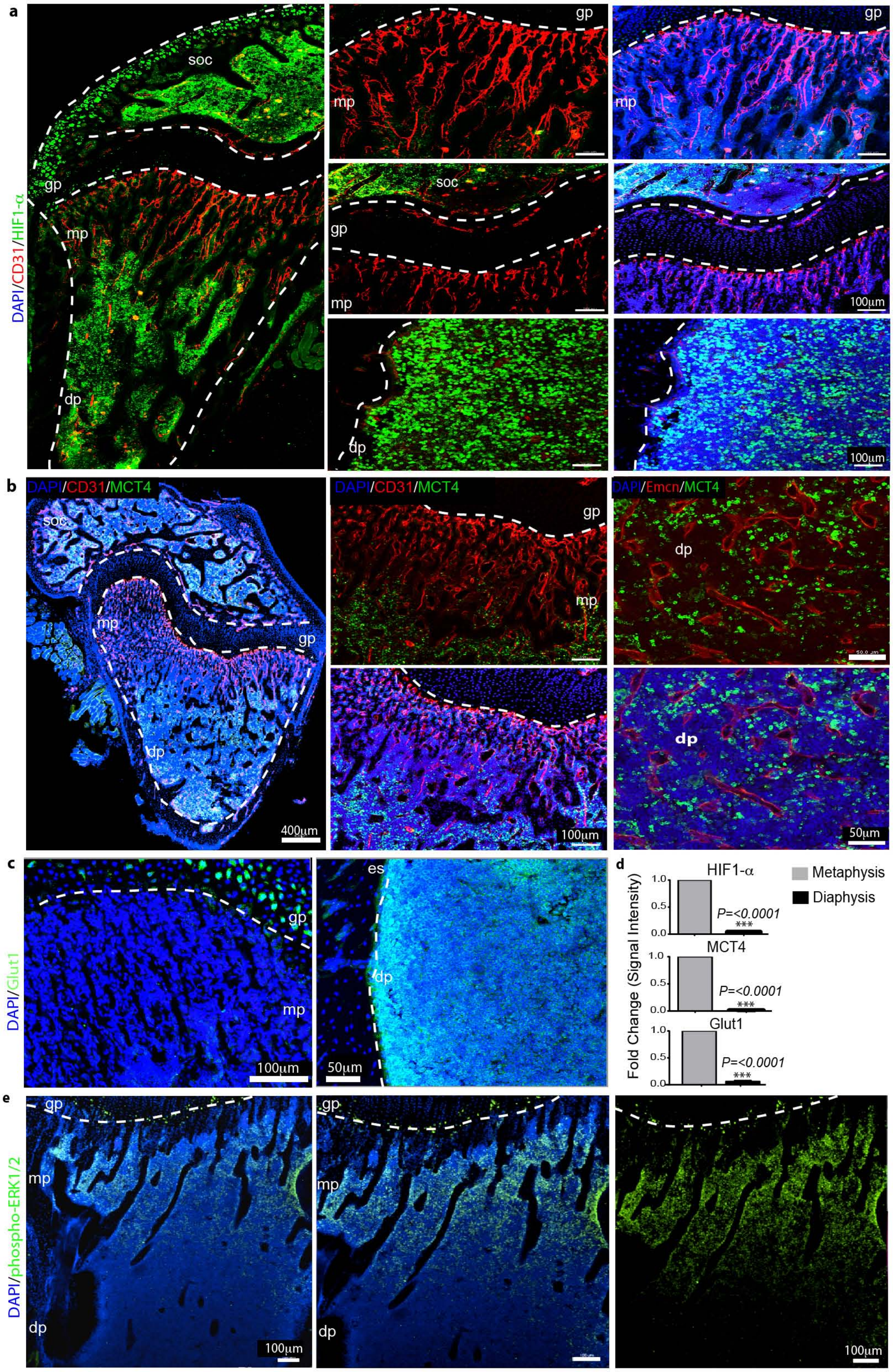




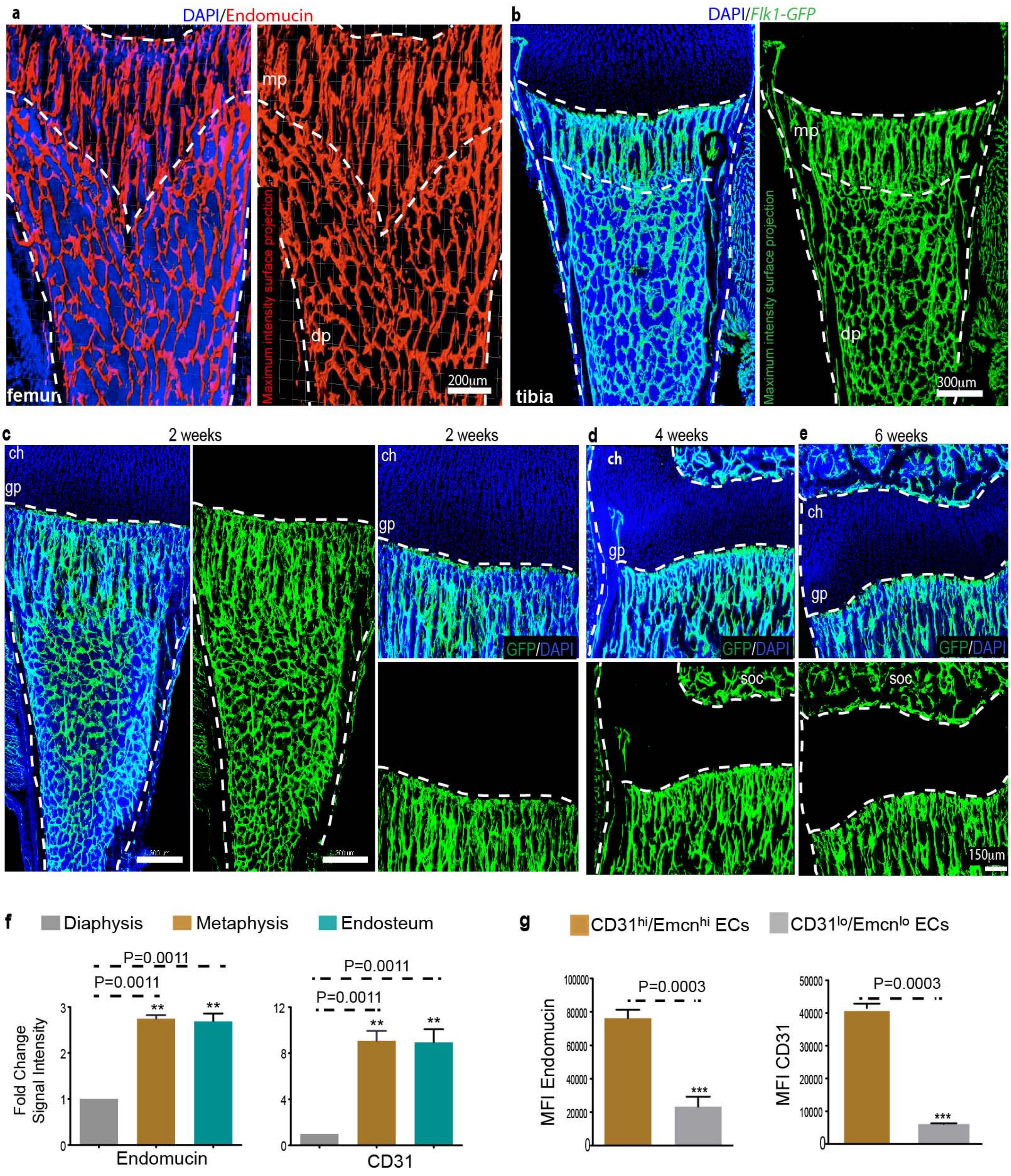
Extended Data Figure 1 | Schematic representation of key findings. a, Type H vasculature (red) in the metaphysis (mp) and endosteum (es) represents a functionally specialized vessel subtype that mediates vessel growth and promotes osteogenesis. The later is presumably mediated by angiocrine growth factors (small circles). In aged animals (right), the number of type H vessels and associated osteoprogenitors (OPs) is strongly reduced so that bone mainly contains type L, sinusoidal vessels characteristic for the diaphyseal (dp) marrow cavity. Arrows indicate the incoming arterial flow and venous drainage. **b**, Regulation of type H endothelium by hypoxia-inducible factor. Endothelial-cell-specific gene inactivation of HIF-1α led to pronounced reduction of type H endothelial cells and osteoprogenitors, whereas the opposite effect was obtained by disrupting endothelial VHL expression. Type H endothelial cells, associated osteoprogenitor cells and bone formation were stimulated by DFM in aged mice.



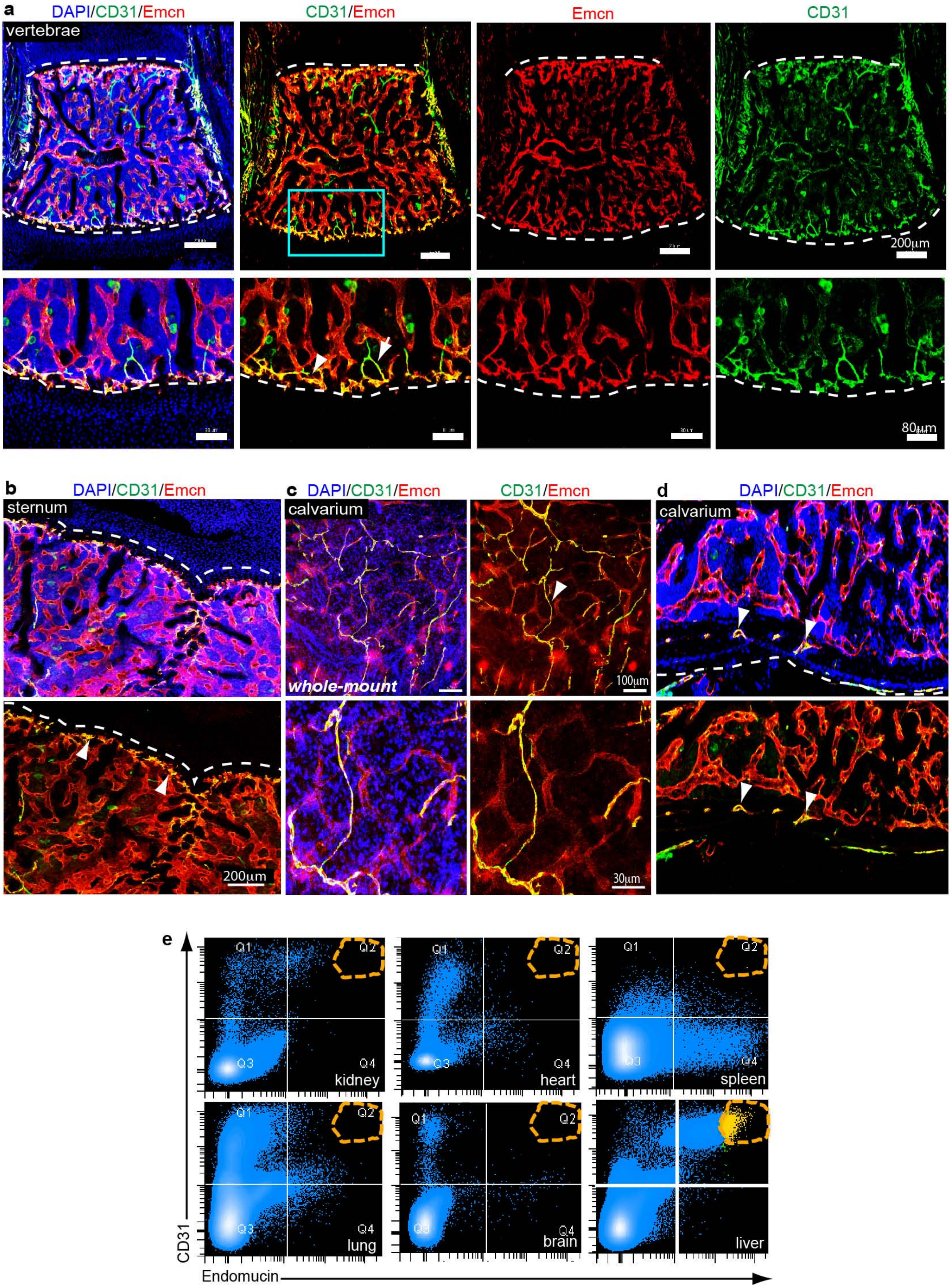
Extended Data Figure 2 | Regional differences in metabolic marker expression. **a**, Tile scan confocal images showing maximum intensity surface projection of Emcn (red) immunostaining on tibial bone section. Nuclei in left image are stained with DAPI (blue). Arrowheads mark the exit of the vein through the cortical bone (cb). Indicated are growth plate (gp), diaphysis (dp), metaphysis (mp) and chondrocytes (ch). **b**, Schematic representation of proximal tibial bone indicating localization of different regions: secondary ossification centre (soc), growth plate (gp), metaphysis (mp), diaphysis (dp), endosteum (es). **c**, Representative confocal images showing pimonidazole (green) staining on a tibial section from a 5-week-old mouse. Nuclei, DAPI (blue); ECs, Endomucin (red, Emcn). Note abundance of pimonidazole staining thoughout the diaphysis (dp) but not in the metaphyseal (mp) region. Dashed lines indicate the borders of cortical bone and growth plate (gp). **d**, **e**, Maximum intensity projections of pimonidazole (green) stained 8-week-old tibia. Nuclei, DAPI (blue); CD31 (red, **d**); CD45 (red, **e**). Green staining is seen in $CD45^{+}$ haematopoietic cells in the diaphysis (dp) and on the bone surface (arrowheads). Dashed lines indicate the borders of cortical bone (cb) and growth plate (gp).



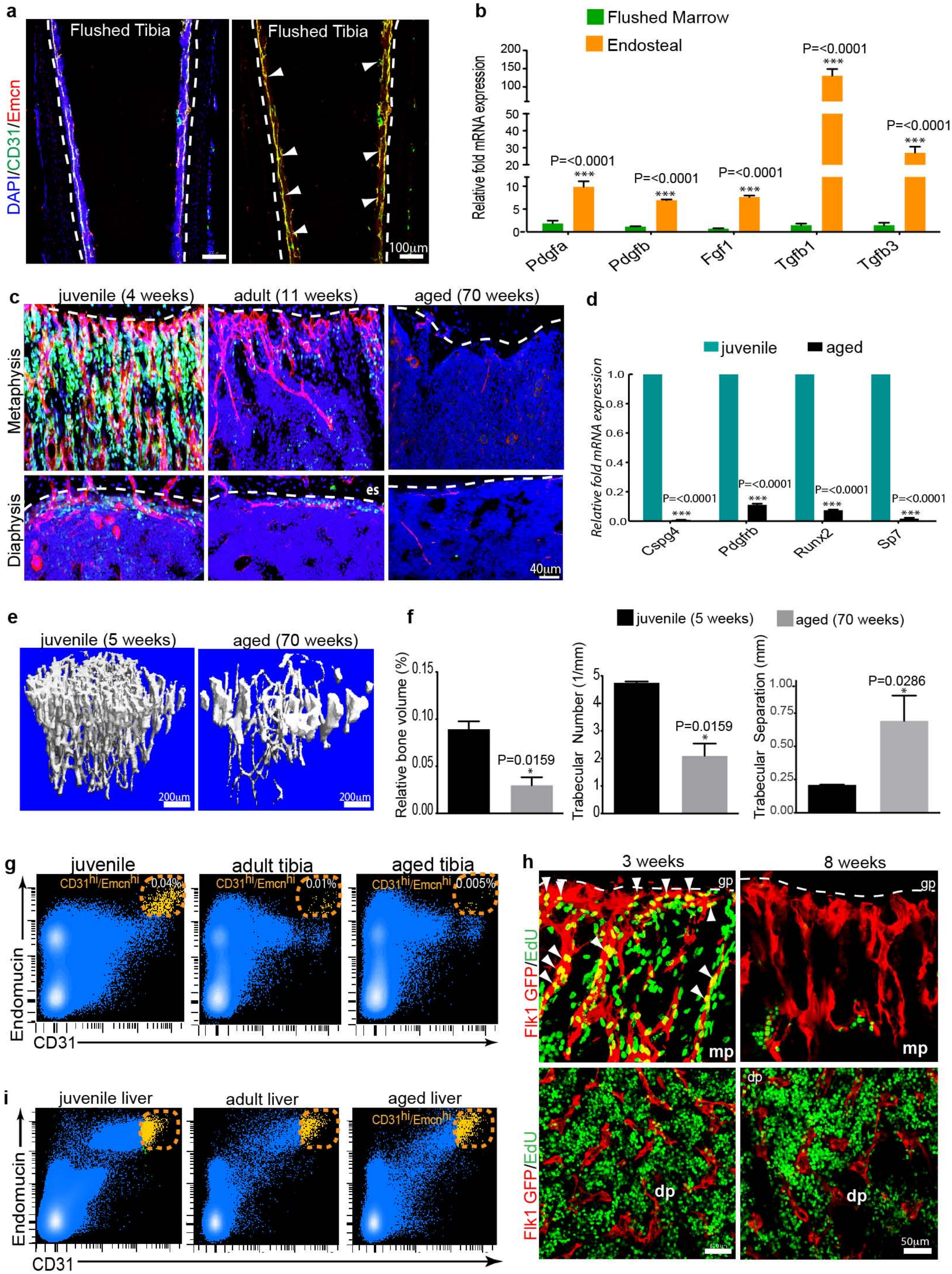
Extended Data Figure 3 | Regional differences in metabolic marker expression. **a**, Representative confocal images showing HIF1-α (green) and CD31 (red) immunostaining on sections of 7-week-old tibiae. Nuclei, DAPI (blue). Note abundance of HIF1α-positive nuclei in the diaphysis (dp) and secondary ossification centre (soc) but not in the metaphyseal (mp) region near the growth plate (gp). Dashed lines indicate borders of the growth plate (top and centre) or, in panels on the bottom, the endosteum (es). **b**, Maximum intensity projections of tibial sections from 7-week-old mice showing immunostaining for the indicated markers. MCT4-positive (green) cells were absent in the metaphysis (mp) but abundant in the diaphysis (dp) and secondary ossification centre (soc). **c**, Maximum intensity projections of Glut1 (green) immunostaining on sections of 7-week-old tibiae. Nuclei, DAPI (blue). Note abundance of Glut1-positive cells in the diaphysis (dp) but not in the metaphyseal (mp) region below the growth plate (gp). Dashed lines indicate borders of growth plate or endosteum (es), respectively. **d**, Quantitation of HIF1-α, MCT4 and Glut1 immunostaining intensities in metaphysis and diaphysis. Data represent mean ± s.e.m. ($n = 5$ mice in two independent experiments). P values, two-tailed unpaired t-test. **e**, Representative tile scan confocal image showing the distribution of phospho-ERK1/2 (green) immunosignal in sections of 7-week-old tibia. Nuclei, DAPI (blue). Note prominent phospho-ERK1/2 staining of cells in the metaphysis (mp) relative to the diaphysis (dp). Dashed line indicates border of the growth plate (gp).



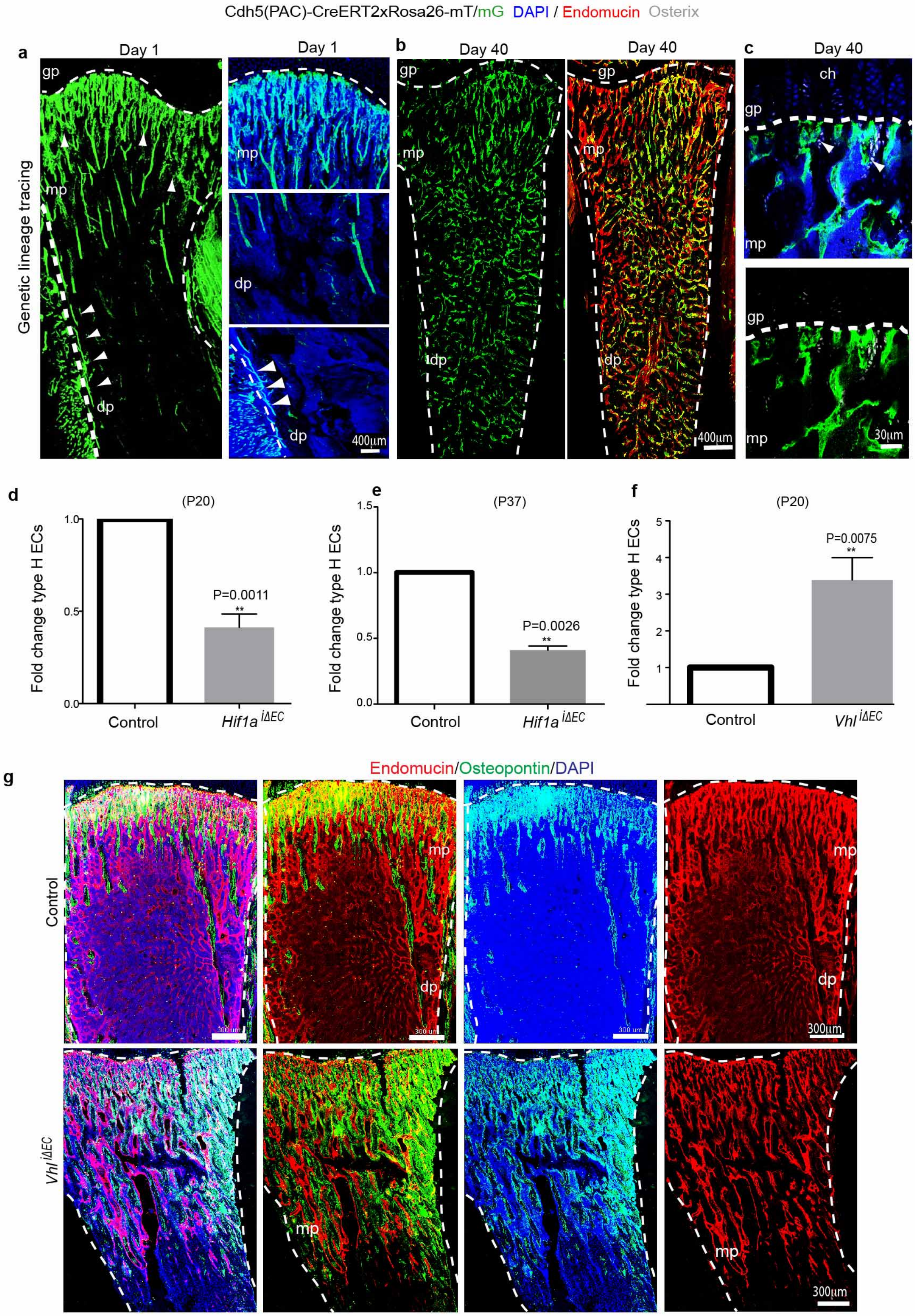
Extended Data Figure 4 | Structural and marker heterogeneity in bone sinusoidal endothelium. **a**, Representative tile scan confocal image showing maximum intensity surface projection of Endomucin (red) immunostaining of ECs in the femur of a 2-week-old mouse. Differences in Endomucin staining intensity are lost in this projection. Nuclei in left image are stained with DAPI (blue). Dashed lines indicate the adjacent growth plate (top), the border of the diaphysis (dp) and the morphologically distinct metaphyseal (mp) and diaphyseal (dp) vessels. **b**, Tile scan confocal image showing maximum intensity surface projection of GFP$^+$ ECs in 2-week-old *Flk1-GFP* transgenic tibia. Left: nuclei, DAPI (blue). Dashed lines indicate the adjacent growth plate (top) or the border of the diaphysis (dp). Dashed lines mark borders of the growth plate (top) and cortical bone (left and right) as well as the interface between column-like metaphyseal (mp) vessels and the highly branched diaphyseal (dp) vasculature. **c–e**, Representative tile scan confocal images of the GFP$^+$ (green) endothelium in *Cdh5(PAC)-CreERT2*, *Rosa26-mT/mG* double transgenic tibiae from 2-week-old (**c**), 4-week-old (**d**), or 6-week-old (**e**) mice after postnatal tamoxifen administration. Nuclei, DAPI (blue). GFP signal is restricted to vessels and absent in chondrocytes of the growth plate (gp) or haematopoietic cells. Dashed lines indicate the borders of the growth plate, cortical bone and secondary ossification centre (soc). **f**, Quantitative analysis of relative CD31 and Endomucin immunostaining intensities in the microvasculature of the metaphysis, diaphysis (marrow cavity) and endosteum, as indicated. Data represent mean ± s.e.m. ($n = 7$ mice from seven independent experiments). *P* values, two-tailed unpaired *t*-test. **g**, Mean fluorescence intensities (MFI) of CD31^{hi}Emcn^{hi} and CD31^{lo}Emcn^{lo} endothelial subsets as determined by flow cytometric analysis of bone marrow cells stained with CD31 and Endomucin. Data represent mean ± s.e.m. ($n = 7$ mice two independent experiments). *P* values, two-tailed unpaired *t*-test.



Extended Data Figure 5 | EC subsets in different skeletal elements and organs. **a–d**, Representative tile scan confocal images showing CD31 (green) and Endomucin (red) immunostaining in juvenile (4-week-old) vertebra (**a**), sternum (**b**), and whole-mount (**c**) or sectioned (**d**) calvarium (parietal bone). Nuclei, DAPI (blue). Arrowheads indicate CD31^{hi}Emcn^{hi} endothelium (yellow). Arrow in **a** marks an adjacent artery (green). **e**, Representative dot plots showing flow cytometric analysis of CD31 and Endomucin-stained single cell suspensions from kidney, heart, spleen, lung, brain and liver. Note absence of a CD31^{hi}Emcn^{hi} EC subset (orange dashed circle in Q2) in these organs with exception of liver.

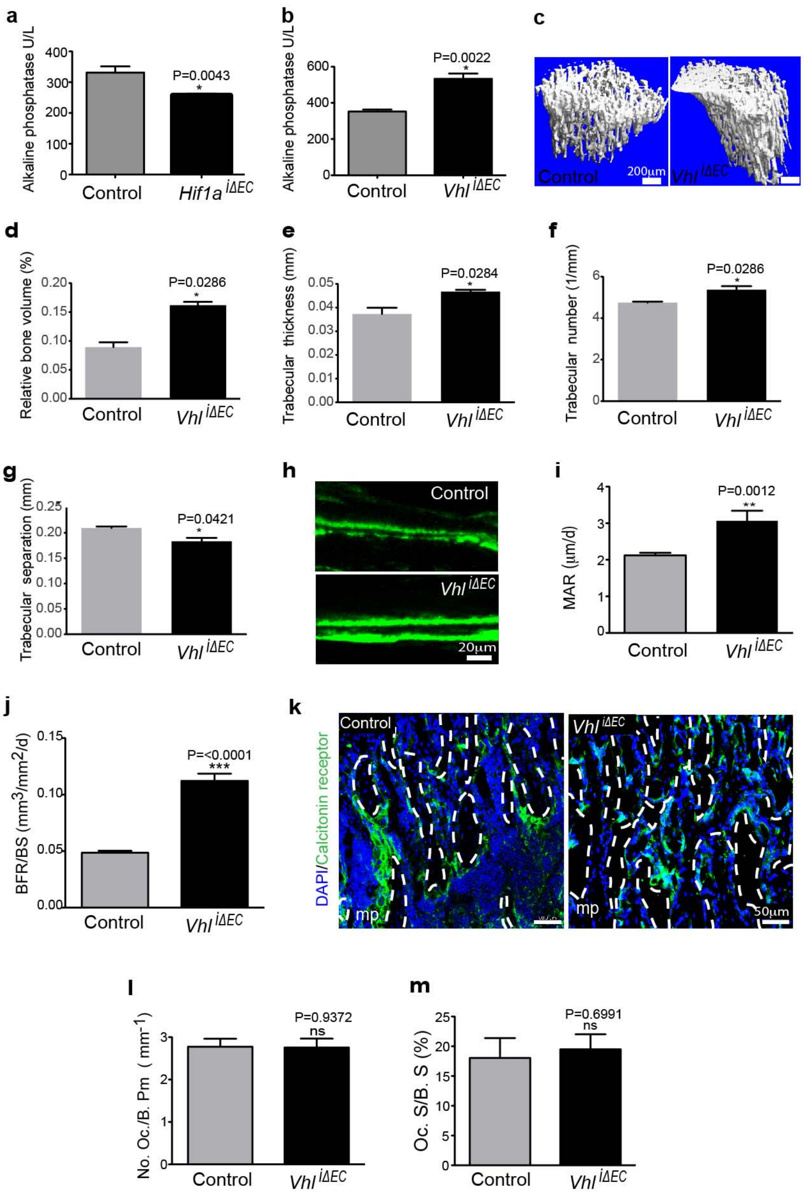

a
Flushed Tibia
Flushed Tibia
DAPI/CD31/Emcn
100μm
b
Flushed Marrow
Endosteal
Relative fold mRNA expression
P=<0.0001
***
Pdgfa
Pdgfb
Fgf1
Tgfb1
Tgfb3
c
juvenile (4 weeks)
adult (11 weeks)
aged (70 weeks)
Metaphysis
Diaphysis
es
40μm
d
juvenile
aged
Relative fold mRNA expression
Cspg4
Pdgfrb
Runx2
Sp7
e
juvenile (5 weeks)
aged (70 weeks)
200μm
f
juvenile (5 weeks)
aged (70 weeks)
Relative bone volume (%)
P=0.0159
Trabecular Number (1/mm)
Trabecular Separation (mm)
P=0.0286
g
juvenile
adult tibia
aged tibia
CD31hi/Emcnhi
0.04%
0.01%
0.005%
Endomucin
CD31
h
3 weeks
8 weeks
gp
mp
dp
Flk1 GFP/EdU
50μm
i
juvenile liver
adult liver
aged liver
CD31hi/Emcnhi
Endomucin
CD31

Extended Data Figure 6 | Regional and age related differences in ECs and bone. **a**, Confocal images showing diaphyseal region of flushed tibia immunostained for CD31 (green) and Endomucin (red). Nuclei, DAPI (blue). Note retention of type H endothelium in the endosteum after flushing. **b**, Comparative qPCR analysis of marrow cell suspension flushed from the tibial diaphysis and compact bone plus endosteum (harbouring type H endothelium). Shown are expression levels of *Pdgfa*, *Pdgfb*, *Fgf1*, *Tgfb1* and *Tgfb3* mRNAs relative to mRNA for β-actin. Data represent mean ± s.e.m. ($n = 6$–8 mice in two independent experiments). *P* values, two-tailed unpaired *t*-test. **c**, Representative confocal images from metaphyseal and diaphyseal regions of tibias from mice of different ages immunostained for Osterix (green) and CD31 (red). Nuclei, DAPI (blue). Dashed lines mark the adjacent growth plate (metaphysis) or endosteum (es) in diaphysis. Note striking decline of $CD31^+$ vessels and associated osteoprogenitors in ageing mice. **d**, Quantitative mRNA expression analysis of *Cspg4*, *Pdgfrb*, *Runx2* and *Sp7* relative to transcripts encoding β-actin in long bones from juvenile and aged mice. Note significant decline of all 4 markers in bone from aged mice. Data represent mean ± s.e.m. ($n = 7$ mice in two independent experiments). *P* values, two-tailed unpaired *t*-test. **e**, Representative μ-CT images of tibias from juvenile (5-week-old) and aged mice. Note significant loss of bone in aged mice. **f**, Quantitative μ-CT analysis of relative bone volume (bone volume/total volume), number of trabeculae and trabecular separation (that is, space between trabeculae) in proximal tibias from juvenile mice and aged mice. Data represent mean ± s.e.m. ($n = 5$ mice in two independent experiments). *P* values, two-tailed unpaired *t*-test. **g**, FACS plots of CD31 and Endomucin double stained single cell suspensions from murine tibias. $CD31^{hi}Emcn^{hi}$ ECs decline with age. **h**, Confocal images showing type H endothelium identified as GFP^+ (red) ECs and proliferation (EdU incorporation, green) in the metaphysis (mp, upper panel) or diaphysis (dp, lower panel) from 3- or 8-week-old *Flk1-GFP* transgenic tibia, as indicated. EdU^+ proliferating GFP^+ cells (arrowheads), which represent type H endothelium, are abundant in the metaphysis (mp) of juvenile mice and but not adult mice. EdU^+GFP^+ ECs are sparse in sinusoidal (type L) vessels of the diaphysis. **i**, FACS plots showing CD31 and endomucin double staining of single cell suspensions from juvenile, adult and aged mice livers. $CD31^{hi}Emcn^{hi}$ ECs in liver do not decline with age.

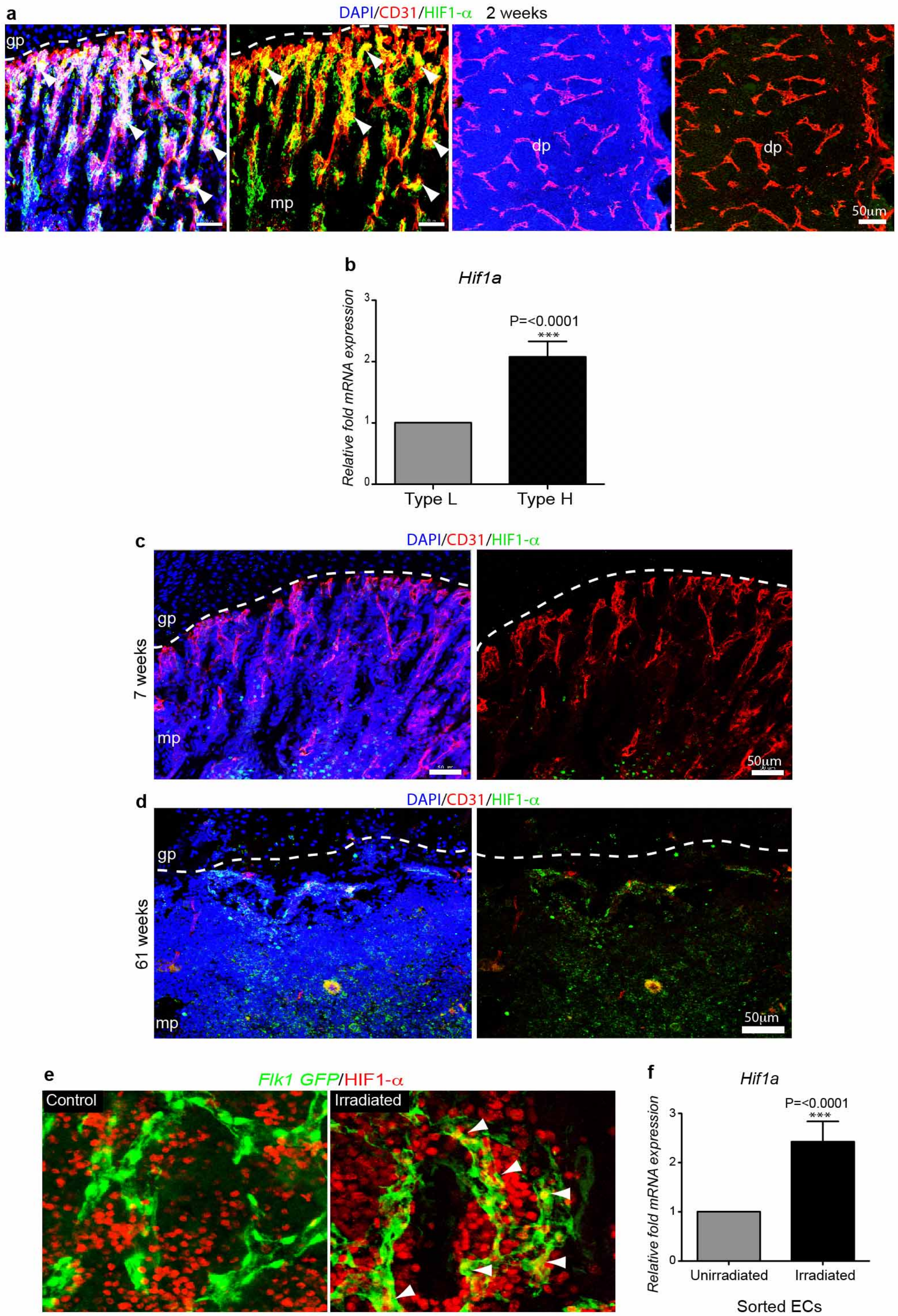


Cdh5(PAC)-CreERT2xRosa26-mT/mG DAPI / Endomucin Osterix
a
Day 1
Day 1
Genetic lineage tracing
gp
mp
dp
400μm
b
Day 40
Day 40
gp
mp
dp
400μm
c
Day 40
ch
gp
mp
30μm
d
(P20)
Fold change type H ECs
1.0
0.5
0.0
P=0.0011
**
Control
Hif1a iΔEC
e
(P37)
Fold change type H ECs
1.5
1.0
0.5
0.0
P=0.0026
**
Control
Hif1a iΔEC
f
(P20)
Fold change type H ECs
5
4
3
2
1
P=0.0075
**
Control
Vhl iΔEC
g
Endomucin/Osteopontin/DAPI
Control
Vhl iΔEC
mp
dp
300μm

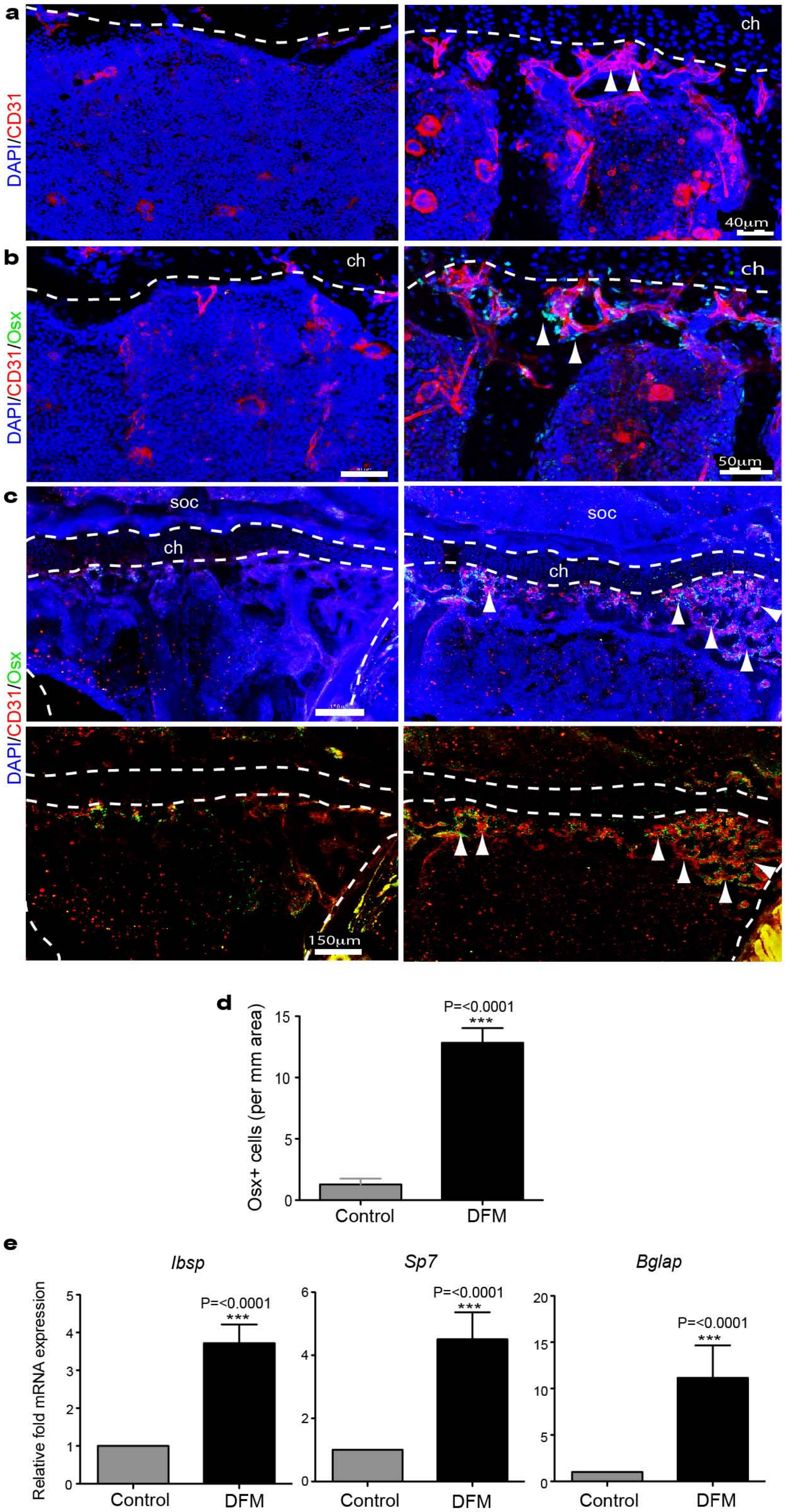
Extended Data Figure 7 | Lineage tracing of type H endothelium and analysis of HIF pathway mutants. **a**, **b**, Representative confocal images of sectioned tibiae from the genetic lineage tracing experiment analysed at 1 and 40 days after tamoxifen administration. GFP-labelled endothelial cells (green) seen in *Cdh5(PAC)-CreERT2*, *Rosa26-mT/mG* double transgenics after 1 day correspond arteries and type H endothelium (arrowheads) in the metaphysis (mp) and endosteum, but not in sinusoidal vessels of the diaphysis (dp). Counterstaining of ECs with anti-Endomucin antibody (red fluorescence) in samples taken after 40 days shows expansion of the GFP^+ endothelium into the diaphyseal microvasculature. Nuclei in small insets in (**a**), DAPI (blue). Dashed lines indicate border of growth plate (gp) and outline of compact bone. **c**, Representative confocal images of metaphyseal (mp) region in tibiae from the genetic lineage tracing experiment analysed at day 40 after tamoxifen administration showing GFP-labelled ECs (green) in *Cdh5(PAC)-CreERT2*, *Rosa26-mT/mG* double transgenics and Osterix staining (white). Nuclei stained with DAPI (blue). Dashed lines drawn below chondrocytes (ch) lines indicate border of growth plate (gp). **d**, **e**, Quantitative analysis of $CD31^{hi}Emcn^{hi}$ endothelial cells in long bone from *Hif1a*$^{i\Delta EC}$ and corresponding littermate controls analysed at postnatal day 20 (P20, **d**) or P37 (**e**). Shown is fold change in frequency of endothelial cells $CD31^{hi}Emcn^{hi}$ ECs identified by flow cytometry. Data represent mean ± s.e.m. ($n = 7$ mice from three independent experiments (**d**); $n = 6$ mice from three independent experiments (**e**)). *P* values, two-tailed unpaired *t*-test. **f**, Quantitative analysis of $CD31^{hi}Emcn^{hi}$ endothelial cells in long bone from *Vhl*$^{i\Delta EC}$ and corresponding littermate controls analysed at postnatal day (P20). Shown is fold change in frequency of endothelial cells $CD31^{hi}Emcn^{hi}$ endothelial cells identified by flow cytometry. Data represent mean ± s.e.m. ($n = 5$ mice from three independent experiments). *P* values, two-tailed unpaired *t*-test. **g**, Representative tile scan confocal images from tibia sections of control and *Vhl*$^{i\Delta EC}$ mutants immunostained for Endomucin (red) and Osteopontin (green). Nuclei, DAPI (blue). Note widespread osteopontin staining in *Vhl*$^{i\Delta EC}$ mutant tibia. gp, growth plate; mp, metaphysis; dp, diaphysis; es, endosteum.



Extended Data Figure 8 | $Vhl^{i\Delta EC}$ mutants show increased bone mass. **a**, **b**, Serum alkaline phosphatase levels in *Hif1a*$^{i\Delta EC}$ (**a**) and $Vhl^{i\Delta EC}$ (**b**) mutants. Data represent mean ± s.e.m. ($n = 5$ or 6 mice for *Hif1a*$^{i\Delta EC}$ mice from three independent experiments; $n = 6$ mice for $Vhl^{i\Delta EC}$ from three independent experiments). *P* values, two-tailed unpaired *t*-test. **c**, Representative μ-CT images of tibias from $Vhl^{i\Delta EC}$ mutants and littermate controls. **d–g**, Quantitative μ-CT analysis of relative bone volume (bone volume/total volume **d**), trabecular thickness (**e**) trabecular number (**f**), and trabecular separation (**g**) in proximal tibia from $Vhl^{i\Delta EC}$ mutants and their littermate controls. Data represent mean ± s.e.m. ($n = 4$ mice from from two independent experiments). *P* values, two-tailed unpaired *t*-test. Note increased bone mass in $Vhl^{i\Delta EC}$ mutants. **h**, Calcein double labelling of 5-week-old $Vhl^{i\Delta EC}$ mutant and littermate control tibiae. **i**, **j**, Quantitative analysis of bone formation parameters. Mineral apposition rate (MAR; **i**) and bone formation rate/bone surface (BFR/BS; **j**) for $Vhl^{i\Delta EC}$ mutants and controls. Data represent mean ± s.e.m. ($n = 6$ or 7 mice from three independent experiments). *P* values, two-tailed unpaired *t*-test. **k**, Representative confocal images showing Calcitonin receptor staining (osteoclasts) in tibia sections from $Vhl^{i\Delta EC}$ mutants and littermate controls. Nuclei, DAPI (blue). **l**, **m**, Histomorphometric analysis of $Vhl^{i\Delta EC}$ and control tibiae showing osteoclast number/bone perimeter (No. Oc./B. Pm; **l**) and osteoclast surface/bone surface (Oc. S/B. S; **m**). Data represent mean ± s.e.m. ($n = 6$ mice from three independent experiments). *P* values, two-tailed unpaired *t*-test.



Extended Data Figure 9 | Age-dependent endothelial HIF1-α expression. **a**, Representative confocal images showing HIF1-α (green) and CD31 (red) immunostaining on sections of 2-week-old tibia. Nuclei, DAPI (blue). Note abundance of HIF1-α-positive type H ECs in 2-week-old metaphysis (mp) but not in the type L endothelium in diaphysis (dp). Dashed line marks border of growth plate (gp). **b**, Quantitative mRNA expression analysis of *Hif1a* transcripts relative to mRNA encoding β-actin in type H and type L. Data represent mean ± s.e.m. ($n = 3$ biological replicates). *P* values, two-tailed unpaired *t*-test. **c, d**, Maximum intensity projections of HIF1-α (green) and CD31 (red) immunostaining in 7-week-old (**c**) and 61-week-old (**d**) tibiae. Nuclei, DAPI (blue). HIF1-α-positive endothelium was not detected in metaphysis (mp) of 7-week-old (**c**) and 61-week-old tibia (**d**). **e**, Maximum intensity confocal images from the diaphysis of 5-week-old *Flk1-GFP* (green) irradiated (900 rads) and control tibiae after HIF1-α (red) immunostaining. HIF1-α signals (arrowheads) in GFP^+ endothelial cells are enhanced after irradiation. **f**, qPCR expression analysis of *Hif1a* relative to transcripts encoding β-actin in FACS-isolated endothelial cells from bones of irradiated mice and untreated controls. Data represent mean ± s.e.m. ($n = 7$ mice from three independent experiments). *P* values, two-tailed unpaired *t*-test.



Extended Data Figure 10 | DFM induction of type H endothelial cells and osteoprogenitors. a, b, Representative confocal images of CD31 (red, **a**, **b**) or Osterix (green, **b**) stained tibia sections from aged DFM-treated (right) and control (left) mice (60–65-weeks-old). Low intensity projection shows only CD31^{hi} cells. DFM induces CD31^{hi} vessels and Osterix^{+} osteoprogenitors (arrowheads). Chondrocytes, ch. **c**, Tile-scan confocal images of CD31 (red) and Osterix (green, Osx) from metaphysis region of stained tibia sections from aged DFM-treated (right) and control (left) mice. Low intensity projection shows only CD31^{hi} cells. DFM induces CD31^{hi} vessels and Osterix^{+} osteoprogenitors (arrowheads). Nuclei, DAPI (blue). Dashed lines mark growth plate chondrocytes (ch) and outline of compact bone. Arrowheads indicate Osterix^{+} cells in secondary ossification centre (soc) and DFM-treated metaphysis. **d**, Quantitation of Osterix^{+} osteoprogenitor cells in DFM treated and control parietal bones. Data represent mean ± s.e.m. ($n = 6$ mice from mice from two independent experiments). P values, two-tailed unpaired t-test. **e**, qPCR analysis of *Ibsp*, *Sp7*, *Bglap* and mRNA expression levels relative to *Actb* in the aged DFM or vehicle-treated (Control) parietal bones, as indicated. Data represent mean ± s.e.m. ($n = 6$ or 7 mice from two independent experiments). P values, two-tailed unpaired t-test.

doi:10.1038/nature13147

Unexpected link between an antibiotic, pannexin channels and apoptosis

Ivan K. H. Poon[1,2,3,4], Yu-Hsin Chiu[5], Allison J. Armstrong[1,2,3], Jason M. Kinchen[1,2,3], Ignacio J. Juncadella[1,2,3], Douglas A. Bayliss[5] & Kodi S. Ravichandran[1,2,3]

Plasma membrane pannexin 1 channels (PANX1) release nucleotide find-me signals from apoptotic cells to attract phagocytes. Here we show that the quinolone antibiotic trovafloxacin is a novel PANX1 inhibitor, by using a small-molecule screen. Although quinolones are widely used to treat bacterial infections, some quinolones have unexplained side effects, including deaths among children. PANX1 is a direct target of trovafloxacin at drug concentrations seen in human plasma, and its inhibition led to dysregulated fragmentation of apoptotic cells. Genetic loss of PANX1 phenocopied trovafloxacin effects, revealing a non-redundant role for pannexin channels in regulating cellular disassembly during apoptosis. Increase in drug-resistant bacteria worldwide and the dearth of new antibiotics is a major human health challenge. Comparing different quinolone antibiotics suggests that certain structural features may contribute to PANX1 blockade. These data identify a novel linkage between an antibiotic, pannexin channels and cellular integrity, and suggest that re-engineering certain quinolones might help develop newer antibacterials.

Pannexins are four-pass transmembrane channels identified as a new family of channels for small molecules (up to ~1 kDa) across the plasma membrane[1,2]. Among the three vertebrate pannexin family members (PANX1, PANX2 and PANX3), PANX1 is the most widely expressed[1] and is implicated in regulating neutrophil activation[3], airway inflammation[4], human immunodeficiency virus infection[5], vasoconstriction[6], migraine[7] and other neurological disorders[8,9]. This broad and diverse range of functions may in part arise from pannexin channel-mediated release of purines such as ATP into the extracellular space, where purinergic signalling can influence multiple physiological processes[10,11]. Thus, PANX1 is an attractive therapeutic target for human diseases and we sought to identify small molecules that can modulate PANX1 function.

Trovafloxacin inhibits PANX1 channel activity

Caspase-mediated cleavage of PANX1 carboxy terminus during apoptosis leads to PANX1 channel opening and release of nucleotide find-me signals from early apoptotic cells to recruit phagocytes[12–15] (Fig. 1a). This channel opening also allows the entry of fluorescent dyes, including TO-PRO-3[13,15] (Fig. 1a). We optimized TO-PRO3 uptake by apoptotic Jurkat cells as a reliable, medium-throughput, flow-cytometry-based assay for monitoring PANX1 activity. We tested a 'library of pharmacologically active compounds' ($LOPAC^{1280}$) containing 1,280 small molecules targeting a diverse range of cellular processes—including currently marketed drugs, failed candidates, and bioactive molecules with known activities. The initial screen revealed three potential PANX1 inhibitors that were tested in secondary screens. Among them, trovafloxacin (a quinolone-based antibiotic) was identified as a potent inhibitor of TO-PRO-3 uptake by apoptotic cells (Fig. 1b). The use of trovafloxacin in patients has been linked to serious adverse side effects, including effects on the central nervous system, hepatic toxicity and in some cases mortality, but the molecular target(s) of trovafloxacin in mammalian cells is unclear[16,17]. Trovafloxacin inhibition of PANX1 was dose-dependent, and comparable to the known pannexin inhibitor carbenoxolone (CBX) (Fig. 1c). Trovafloxacin also inhibited ATP release from apoptotic cells (Fig. 1d). Importantly, trovafloxacin did not inhibit caspase 3/7 activation, or caspase-mediated PANX1 cleavage during apoptosis (Extended Data Fig. 1a, b), ruling these out as reasons.

Several additional analyses suggested trovafloxacin could directly target PANX1 channel activity. Adding trovafloxacin to cells already undergoing apoptosis (that is, with open PANX1 channels) acutely blocked TO-PRO-3 uptake (Extended Data Fig. 1c, d). When we measured apoptosis-induced plasma membrane PANX1 currents at the single-cell level, by whole-cell patch-clamp recordings, trovafloxacin rapidly inhibited the inward current (at −50 mV), with minimal effect on outward current (at +80 mV) (Fig. 1e and Extended Data Fig. 1e). We have previously shown that the C-terminal tail of PANX1 blocks the channel pore, and that adding excess soluble C-terminal tails can inhibit 'open' PANX1 channels, especially the inward current (analogous to trovafloxacin)[14]. In contrast, CBX blocked both inward and outward currents[13,18,19] (Fig. 1e, f). Trovafloxacin did not inhibit connexin 43 gap junction or PANX2 (Extended Data Fig. 1f–i). Using a tobacco etch virus (TEV)-protease system to cleave the C-terminal tail of recombinant PANX1 and induce channel activity (independent of apoptosis)[13,14], trovafloxacin again potently blocked open PANX1 channels (Fig. 1g). To test direct channel blocking, we recorded TEV-cleaved PANX1 single-channel activity in excised inside-out patch-clamp by adding trovafloxacin to the patch; this led to an increase in the time spent in the closed state, with open probability (NPo) of ~0.85 in control conditions reduced to ~0.15 with trovafloxacin (Fig. 1h). The half maximal inhibitory concentration (IC_{50}) of trovafloxacin was ~4 μM for the PANX1 inward current (Fig. 1i), similar to concentrations normally achieved in human plasma (2–10 μM)[20,21]. These data indicated that mammalian PANX1 channels could be a direct target of antibiotic trovafloxacin.

Trovafloxacin promotes fragmentation of apoptotic cells

Next, we investigated trovafloxacin effects on apoptotic cells via microscopy and made several surprising observations. Apart from reducing

[1]The Center for Cell Clearance, University of Virginia, Charlottesville, Virginia 22908, USA. [2]Department of Microbiology, Immunology, and Cancer Biology, University of Virginia, Charlottesville, Virginia 22908, USA. [3]Beirne B. Carter Center for Immunology Research, University of Virginia, Charlottesville, Virginia 22908, USA. [4]La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria 3086, Australia. [5]Department of Pharmacology, University of Virginia, Charlottesville, Virginia 22908, USA.

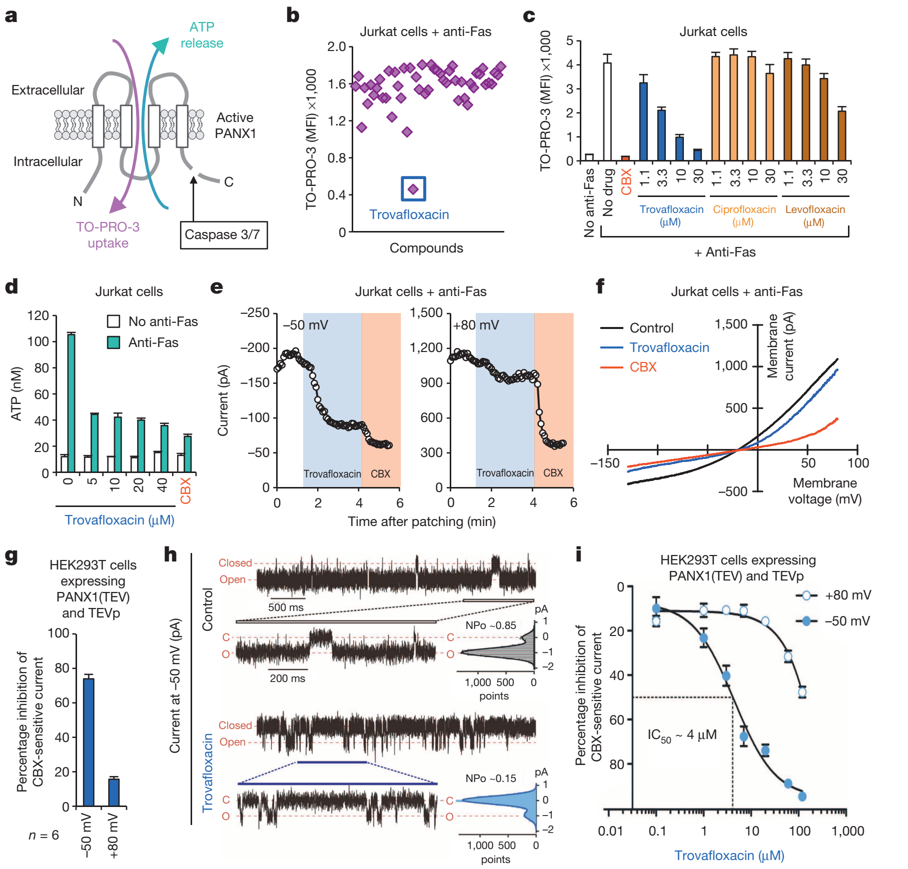


Figure 1 | Trovafloxacin inhibits pannexin 1 activity during apoptosis. **a**, Schematic of PANX1 with ATP release and TO-PRO-3 uptake by apoptotic cells after caspase-mediated channel opening. **b**, Subset of LOPAC^{1280} library screen in modulating TO-PRO-3 uptake (median fluorescence intensity, MFI; average of 1,200–1,800), with inhibition by trovafloxacin indicated. **c**, Dose-dependent inhibition of TO-PRO-3 uptake by trovafloxacin (or positive control CBX), levofloxacin and ciprofloxacin ($n = 3$). **d**, Inhibition of ATP release from apoptotic cells by trovafloxacin ($n = 3$). **e**, **f**, Jurkat cells after 2 h with anti-Fas used for patch-clamp recordings. Patch-clamp recordings from apoptotic cells at −50 mV and +80 mV with trovafloxacin (blue shading) or CBX (pink); traces are representative of recordings from 7 cells per group (**e**). Current–voltage relationships (I–V curves) from apoptotic cells over a range of voltages ($n = 7$) (**f**). **g**, Trovafloxacin inhibition of CBX-sensitive current in HEK293T cells expressing TEV-protease activated PANX1 channels ($n = 6$). **h**, Inside-out patch-clamp recordings from excised patches containing TEV-cleaved PANX1 channels showing traces of inward currents at −50 mV. PANX1 channels are mostly 'open' without trovafloxacin (open probability, NPo = 0.85), but mostly closed with trovafloxacin (NPo = 0.15); histograms of channel activity are shown for each trace ($n = 3$). **i**, Inhibition of TEV-cleaved open PANX1 channels by trovafloxacin, with IC_{50} of 4 μM at −50 mV ($n \geq 6$ per data point). Error bars represent s.e.m.

TO-PRO-3 uptake by apoptotic cells, trovafloxacin also induced the formation of smaller particles; these 'fragments of apoptotic cells' were annexin V$^+$, indicating phosphatidylserine exposure, and resembled apoptotic bodies that arise after cell disassembly during apoptosis[22,23] (Fig. 2a, b). To quantify these apoptotic bodies, we designed a flow-cytometry-based assay (Extended Data Fig. 2a, b) to analyse five parameters simultaneously: cell size (forward scatter, FSC), cellular complexity (side scatter, SSC), annexin V staining (indicating apoptosis), TO-PRO-3 uptake (PANX1 opening), and 7-aminoactinomycin D (7-AAD) staining (loss of plasma membrane integrity). We also performed Image-Stream analysis, which combines flow cytometry and image acquisition to confirm the categorization of cells and apoptotic bodies (Extended Data Fig. 2c). The apoptotic bodies were annexin V$^{\text{intermediate}}$ (compared to annexin V$^{\text{high}}$ apoptotic cells), with smaller size and reduced complexity (Extended Data Fig. 2). Trovafloxacin promoted the formation of apoptotic bodies in a concentration-dependent manner (Fig. 2c, d). Importantly, without an apoptotic stimulus, trovafloxacin did not induce apoptosis or apoptotic bodies (Extended Data Fig. 1a), indicating that trovafloxacin probably modulates steps during apoptosis that promote cellular disassembly. Although trovafloxacin itself did not induce apoptosis, because the human body turns over more than 200 billion cells daily via apoptosis as part of homeostasis[24], trovafloxacin could affect these apoptotic steps.

PANX1 regulates the disassembly of apoptotic cells

Although some studies have addressed apoptotic cell disassembly, the mechanisms that control the formation of apoptotic bodies are not well understood[22,23,25,26]. Several lines of evidence suggested a key role for PANX1 in regulating apoptotic cell integrity/disassembly. First, blocking pannexin channels in apoptotic Jurkat cells with CBX or probenecid (another pannexin inhibitor[27]) promoted formation of apoptotic bodies (Fig. 2e, f). Second, this effect was not dependent on the mode of apoptosis induction, as blocking PANX1 during Fas- or ultraviolet-mediated apoptosis promoted formation of apoptotic bodies (Fig. 2f and Extended Data Fig. 3a). Third, using loss of cellular complexity during apoptosis as an indication of fragmentation of larger apoptotic cells (SSC$^{\text{high}}$/higher cellular complexity) into smaller apoptotic bodies (SSC$^{\text{low}}$/lower cellular complexity), blocking PANX1 reduced the number of SSC$^{\text{high}}$ particles (Fig. 2g). Treating wild-type primary thymocytes undergoing apoptosis *ex vivo* with trovafloxacin, CBX or probenecid also led to more apoptotic bodies with reduced cellular complexity (Fig. 3a, b). Fourth, the enhanced apoptotic bodies correlated inversely with TO-PRO-3 uptake (Extended Data Fig. 3b), with progressively more PANX1 channel blocking leading to greater formation of apoptotic bodies. Fifth, in a genetic approach, overexpressing a dominant negative form of PANX1 (PANX1 DN) with a mutant caspase cleavage site that can interfere with endogenous PANX1 function[13] (Extended Data Fig. 4a) resulted in increased formation of apoptotic bodies (Fig. 2h). Conversely, overexpressing wild-type PANX1 led to fewer apoptotic bodies. This 'gain of function' due to wild-type PANX1 expression was still dependent on PANX1 channel activity, with reversal by PANX1 blockers (Fig. 2i). Lastly, blocking PANX1 channel functions did not affect some known apoptotic features such as DNA fragmentation (Extended Data Fig. 4b) or phosphatidylserine exposure (Fig. 2b). Collectively, these data identify PANX1 as a regulator of cellular disassembly during apoptosis.

Trovafloxacin and PANX1 regulate apoptotic cell disassembly *in vivo*

To address how trovafloxacin and PANX1 might affect apoptosis in a whole animal, we used intraperitoneal dexamethasone injection to induce synchronous apoptosis of a large fraction of mouse thymocytes *in vivo*[12,28] (Fig. 3c). Administering trovafloxacin with dexamethasone promoted the fragmentation of the apoptotic thymocytes (Fig. 3d, with fewer SSC$^{\text{high}}$ particles; see Extended Data Figs 5 and 6 for gating strategy). However, in these experiments we could not track thymic trovafloxacin concentrations, and trovafloxacin *in vivo* could also have had other effects beyond PANX1. Therefore, we generated mice carrying deletion

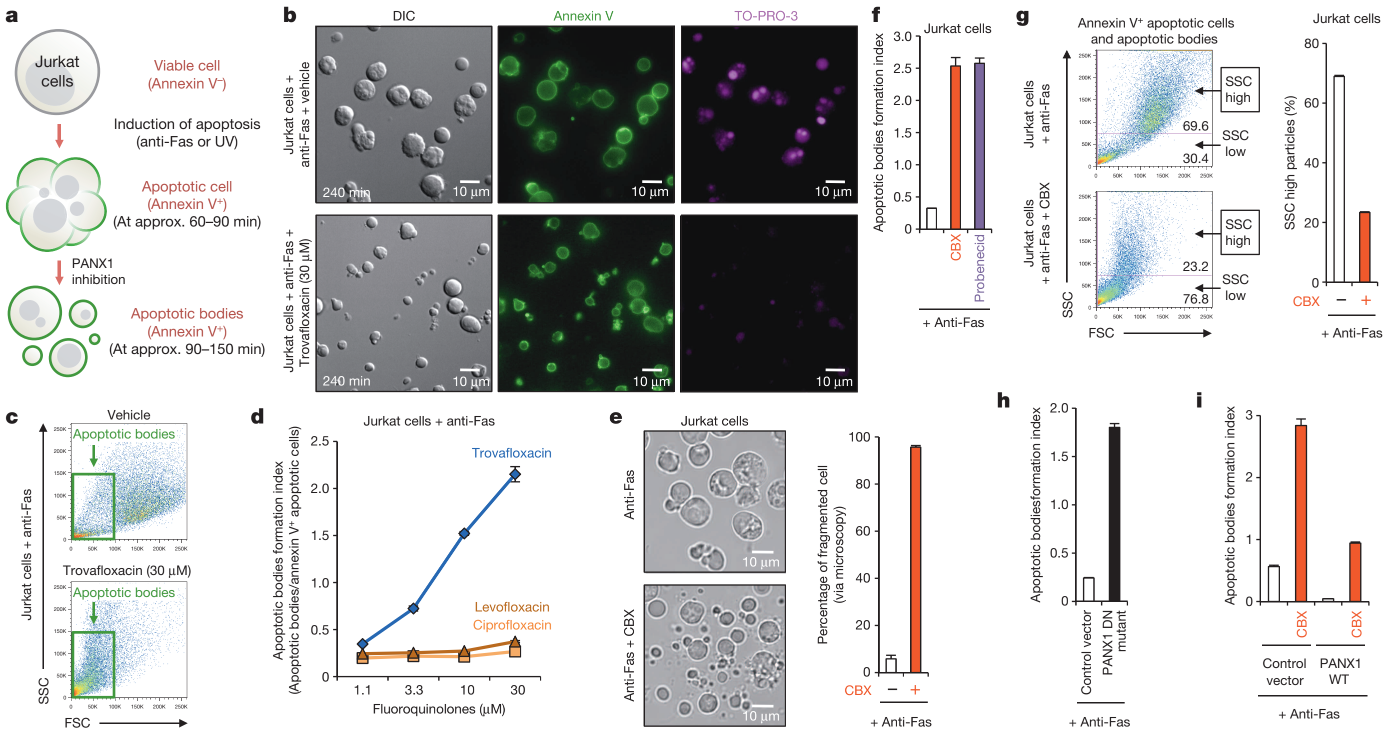


Figure 2 | Trovafloxacin-mediated inhibition of PANX1 promotes formation of smaller apoptotic bodies. **a**, Schematic for apoptosis progression and generation of apoptotic bodies, derived from multi-parameter time-lapse analyses of Jurkat cells. **b**, Images of vehicle- or trovafloxacin-treated apoptotic cells stained with annexin V or TO-PRO-3. Note the annexin $V^{positive}$, TO-PRO-$3^{negative}$ apoptotic bodies (∼0.9–5 μm) after trovafloxacin treatment. **c**, Trovafloxacin increases particles with reduced size (FSC) and complexity (SSC). **d**, Generation of apoptotic bodies is enhanced by trovafloxacin, but not levofloxacin and ciprofloxacin ($n = 3$). **e**, Left, representative 4 h time-lapse images monitoring apoptotic cell morphology (454 cells for untreated and 355 cells for CBX-treated, $n = 3$). Right, percentage of cell fragmentation (≥2 fragments). **f**, Apoptotic bodies formation index (apoptotic bodies/annexin V^+ apoptotic cells) is enhanced by PANX1 blockers CBX and probenecid ($n = 3$). **g**, Left, size (FSC) and complexity (SSC) of particles from apoptotic cells with or without CBX. Right, percentage of SSC^{high} particles. **h**, **i**, Formation of apoptotic bodies is increased in Jurkat cells expressing PANX1 DN mutant, but decreased in PANX1 WT transfected cells ($n = 3$). The PANX1 WT effect is reversed by CBX. Error bars represent s.e.m.

of *Panx1*, using embryonic stem cells engineered to disrupt the *Panx1* gene (Extended Data Fig. 7). In the thymus of $Panx1^{-/-}$ and $Panx1^{+/-}$ mice treated with dexamethasone, the fragmentation of apoptotic cells was enhanced (Fig. 3e), essentially phenocopying the effect of trovafloxacin treatment in wild-type mice (Fig. 3d). The apoptotic particles from $Panx1^{-/-}$ and $Panx1^{+/-}$ mice also showed reduced TO-PRO-3 uptake (Fig. 3e). To rule out that loss of PANX1 in other tissues contributed to the observed thymic phenotypes in the global knockout mice, we specifically deleted *Panx1* in thymocytes by crossing the $Panx1^{fl/fl}$ mice with Lck-Cre mice, where Cre is expressed in the T-cell lineage (Extended Data Fig. 7). Similar apoptotic fragmentation and loss of TO-PRO-3 uptake were observed when PANX1 was deleted specifically in thymocytes (Fig. 3f), indicating a cell-autonomous effect. *Ex vivo*, apoptotic thymocytes from PANX1 global knockout mice or the Lck-Cre/$Panx1^{fl/fl}$ mice also showed enhanced formation of apoptotic bodies, decreased TO-PRO-3 uptake and reduced cellular complexity (Fig. 3g–i). Collectively, these data indicate that PANX1 critically influences cellular integrity of dying cells, with loss of PANX1 function either genetically or via trovafloxacin treatment promoting disassembly of apoptotic cells.

One possible mechanism for this increased cellular fragmentation is the lack of ATP that would normally be released by apoptotic cells, and the various autocrine and paracrine signalling via P2 purinergic receptors[3,5,12,13,29]. However, adding exogenous ATP to apoptotic cells when PANX1 channel activity was blocked did not inhibit apoptotic bodies formation (Extended Data Fig. 3c, d). Conversely, adding recombinant apyrase (to hydrolyse ATP) during induction of apoptosis did not promote the formation of apoptotic bodies (Extended Data Fig. 3e). Furthermore, treatment with suramin, a broad inhibitor of P2 family G-protein-coupled receptors, did not enhance formation of apoptotic bodies (Extended Data Fig. 3f). Thus, the extracellular ATP levels and P2 purinergic receptor signalling are probably not directly involved in regulating apoptotic cell integrity.

PANX1 regulates formation of string-like apoptopodia

We next used time-lapse microscopy to monitor cell morphology and TO-PRO-3 uptake. In our experimental conditions with Jurkat cells induced to undergo Fas-mediated apoptosis, most cells begin to show rounding and membrane blebs around 60–90 min, and continue to bleb for another 30–60 min, a time period referred here as 'dynamic blebbing'. Remarkably, very few of these blebs detach from the apoptotic cell, and after cessation of the dynamic blebbing, these blebs remain attached to the dying cell (Fig. 4a, b and Supplementary Video 1). When we added TO-PRO-3 into the medium and tracked its uptake as a measure of PANX1 activity, TO-PRO-3 staining was detected before cell rounding and the onset of blebbing (Fig. 4c and Extended Data Fig. 4c). This indicates that PANX1 channels are activated/opened before detection of membrane blebbing. When we compared cells with normal or impaired PANX1 function (treated with trovafloxacin or CBX, or expressing PANX1 DN mutant), a previously unappreciated step during dynamic membrane blebbing regulated by pannexin channels was revealed. Specifically, apoptotic cells with impaired PANX1 function produced long string-like structures, which we preliminarily denote 'apoptopodia', with blebs of different sizes attached to the ends (Fig. 4a, b, d and Supplementary Videos 2 and 3). These apoptopodia were annexin V^+, indicating phosphatidylserine exposure (Fig. 4e). Whereas few apoptopodia were seen in apoptotic cells with normal PANX1 function, their frequency was substantially increased by blocking PANX1 (Fig. 4f). Blocking PANX1 function in live cells (either pharmacologically or using the PANX1 DN mutant) did not cause apoptopodia (data not

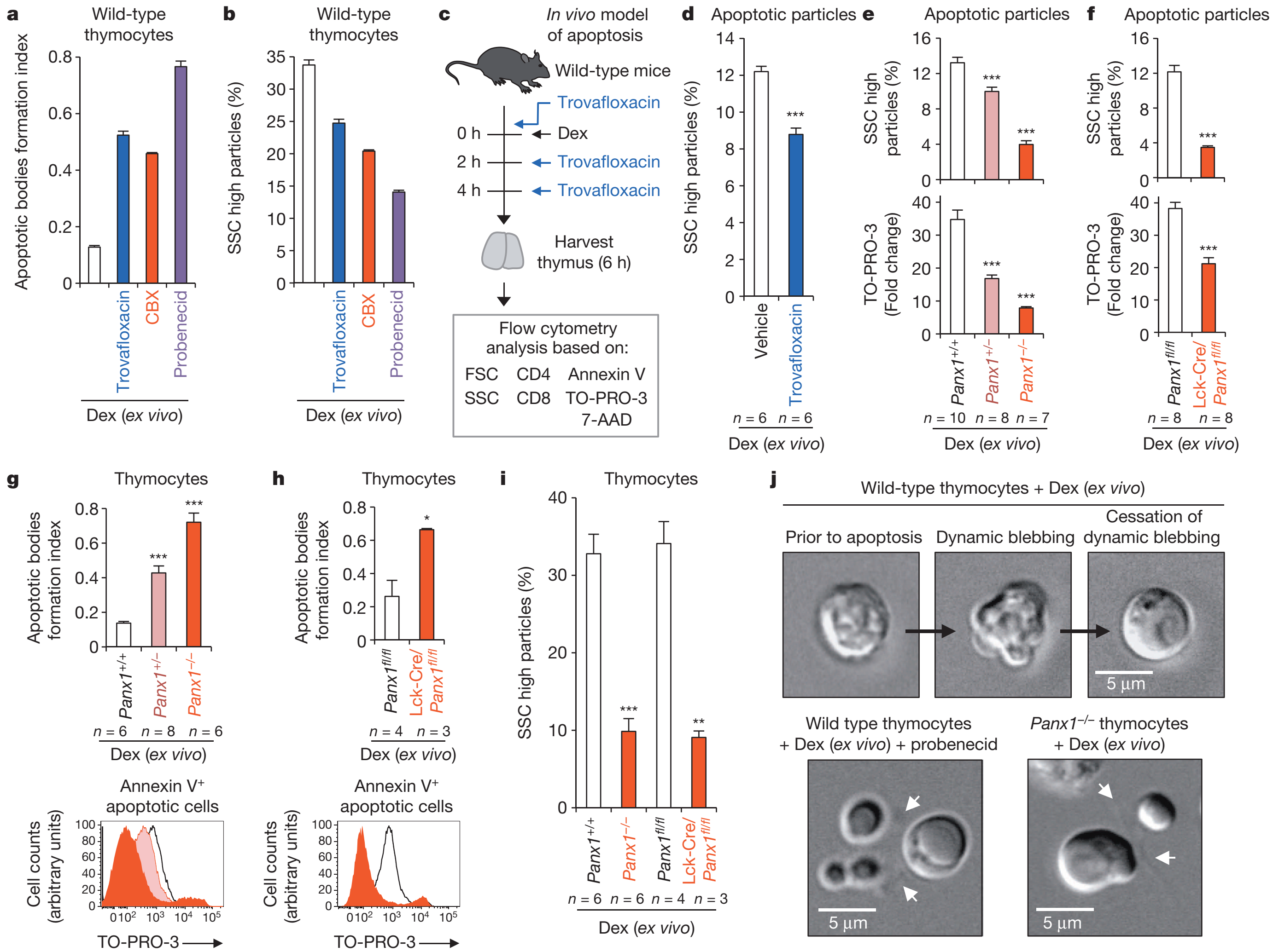


Figure 3 | Pannexin 1 regulates disassembly of apoptotic thymocytes. **a**, **b**, Increase in formation of apoptotic bodies (**a**), or decrease in SSC^{high} particles (**b**) from dexamethasone (Dex)-treated thymocytes *ex vivo* by trovafloxacin, CBX, and probenecid ($n = 3$). **c**, Schematic for testing trovafloxacin in thymic apoptosis *in vivo*. **d**, Reduction in SSC^{high} apoptotic bodies in thymi of wild-type mice treated with Dex+trovafloxacin ($n = 6$). **e**, **f**, Complexity and TO-PRO-3 uptake (fold change relative to unstained) in thymi of indicated mice injected intraperitoneally with Dex ($n = 10$ for *Panx1*$^{+/+}$, $n = 8$ for *Panx1*$^{+/-}$, $n = 7$ for *Panx1*$^{-/-}$, $n = 8$ for *Panx1*$^{fl/fl}$ and Lck-Cre/*Panx1*$^{fl/fl}$). **g**, **h**, Formation of apoptotic bodies (upper) or TO-PRO-3 uptake (lower) *ex vivo* in Dex-treated thymocytes from *Panx1*$^{-/-}$ or Lck-Cre/*Panx1*$^{fl/fl}$ thymocytes. **i**, Reduction in SSC^{high} subcellular apoptotic particles in thymocytes from *Panx1*$^{-/-}$ and Lck-Cre/*Panx1*$^{fl/fl}$ mice. (**g**–**i**, $n = 6$ for *Panx1*$^{+/+}$, $n = 8$ for *Panx1*$^{+/-}$, $n = 6$ for *Panx1*$^{-/-}$, $n = 4$ for *Panx1*$^{fl/fl}$, $n = 3$ for Lck-Cre/*Panx1*$^{fl/fl}$.) **j**, Time-lapse images monitoring apoptotic cell morphology of thymocytes. Arrows, apoptopodia. Error bars represent s.e.m. Data in **a**, **b** and **j** are representative of at least two independent experiments. See Extended Data Figs 5 and 6 for gating strategy. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, unpaired Student's two-tailed *t*-test.

shown), indicating that pannexin channels regulate the formation of these membrane protrusions after apoptosis induction. These apoptopodia were also observed in apoptotic thymocytes from PANX1 knockout mice, or wild-type thymocytes treated with pannexin inhibitors (Fig. 3j). Fas-, dexamethasone- or ultraviolet-induced apoptosis led to apoptopodia formation when PANX1 function was impaired (data not shown).

We next asked how the apoptopodia formation and membrane blebbing were linked. PANX1-mediated TO-PRO-3 uptake began before the onset of blebbing (Fig. 4c), but the apoptopodia were detectable only after the onset of blebbing (Fig. 4b, g). This indicates that PANX1 activity is probably continuously required during apoptosis progression to downregulate the formation of apoptopodia. Interestingly, when blebbing was blocked by inhibiting actomyosin contraction or the Rho-dependent kinase ROCK (which are needed for apoptotic blebbing[22,30,31]), apoptopodia formation was unaffected, with noticeable absence of blebs at the end of these strings (Extended Data Fig. 8a–c). Importantly, inhibiting blebbing reduced the formation of apoptotic bodies in Jurkat cells with PANX1 function disrupted by pharmacological block, or in PANX1-deficient thymocytes (Extended Data Fig. 8d–f). These data indicate that the apoptopodia that arise owing to PANX1 inhibition are independent of blebbing, but that the annexin V$^+$ vesicular structures seen at the end of apoptopodia are dependent on blebbing and contribute to apoptotic bodies formation. LR73 fibroblasts induced to undergo apoptosis also showed the formation of apoptopodia and apoptotic bodies when PANX1 function was impaired (Extended Data Fig. 9). Thus, trovafloxacin-mediated enhancement of apoptotic bodies formation helped uncover a new step during the early stages of apoptosis; specifically, pannexin channel function is required to restrain detachment of blebs from the apoptotic cells, and in turn, regulate the nature of apoptotic cell disassembly (Extended Data Fig. 10).

PANX1 inhibition varies among different quinolones

Although quinolone antibiotics are effective at killing bacteria by targeting their topoisomerases, and are used worldwide to treat various bacteria infections[32–34], some quinolones have serious side effects (incidence up to 5%) in the liver, skin, tendon, gastrointestinal tract, central nervous system and cardiovascular system that are not readily explained[16,17].

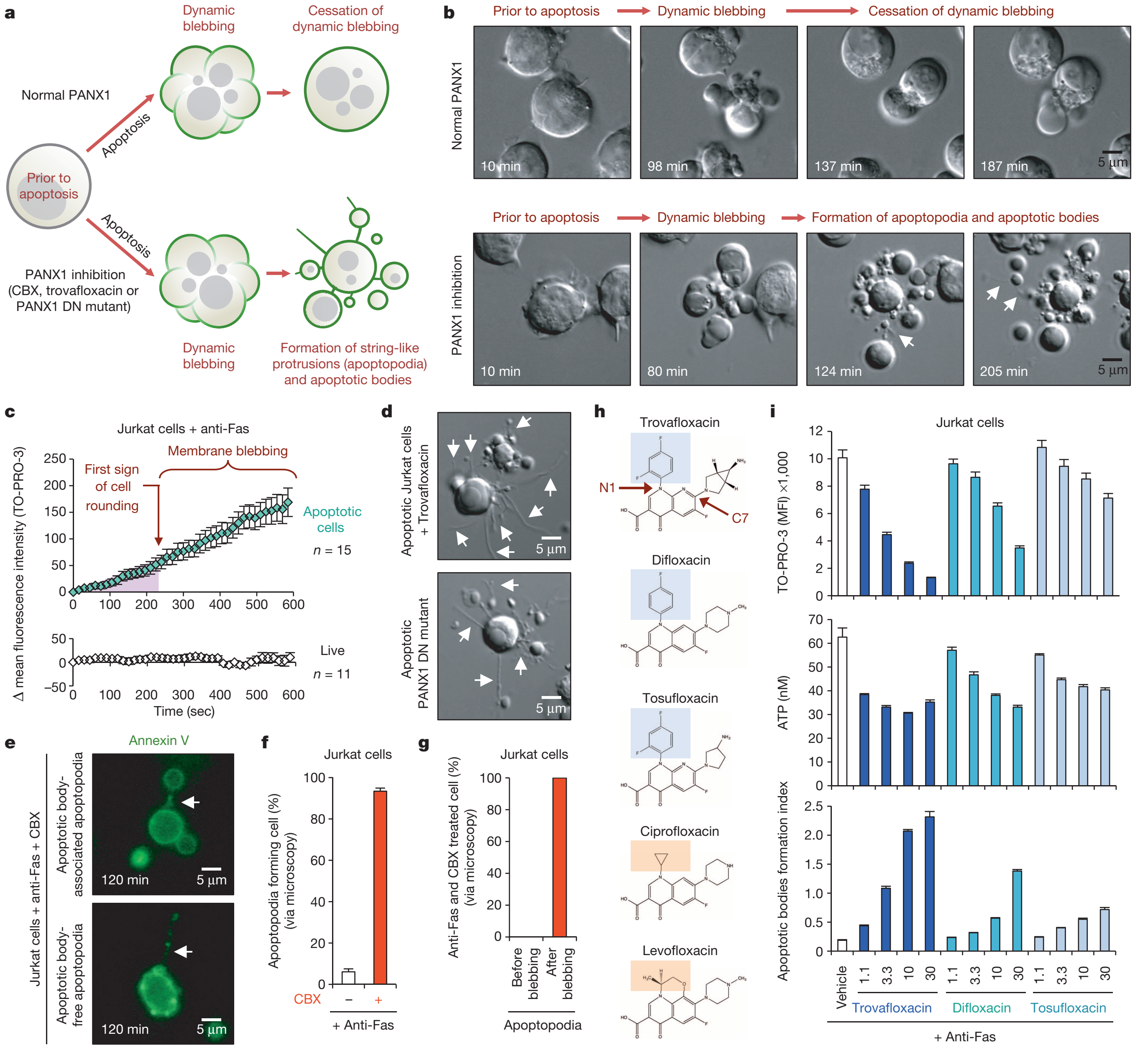


Figure 4 | Formation of string-like apoptopodia after membrane blebbing correlates with formation of apoptotic bodies. **a**, Schematic of apoptotic Jurkat cells with normal or impaired PANX1, based on time-lapse microscopy. **b**, Time-lapse images monitoring progression of apoptotic cell morphology. **c**, PANX1-mediated TO-PRO-3 entry occurs before membrane blebbing as determined by quantifying TO-PRO-3 uptake by apoptotic cells ($n = 15$) in time-lapse imaging (normalized to first sign of cell rounding). TO-PRO-3 uptake was undetectable in live cells ($n = 11$). **d**, Trovafloxacin-treated apoptotic cells or expressing PANX1-DN mutant show formation of apoptopodia. **e**, Apoptopodia have membranes with exposed phosphatidylserine and apoptotic blebs at the end of the protrusion. **f**, Percentage of apoptotic cells with apoptopodia during 4 h of time-lapse imaging. **g**, Apoptopodia are barely detectable before onset of membrane blebbing in cells with impaired PANX1 function, but detected after initiation of blebbing. (**f**, **g**, 429 and 302 cells for untreated and CBX-treated, respectively, $n = 3$.) **h**, Chemical structures of quinolones used, with the N1 fluorinated ring (blue) indicated in quinolones that inhibit PANX1 (trovafloxacin, difloxacin and tosufloxacin), compared to quinolones with no apparent PANX1 inhibition (ciprofloxacin and levofloxacin) (orange). **i**, TO-PRO-3 uptake, ATP release, and formation of apoptotic bodies from apoptotic Jurkat cells treated with trovafloxacin, difloxacin or tosufloxacin ($n = 3$). Error bars represent s.e.m. Arrows, apoptopodia. Data are representative of at least two independent experiments.

Surprisingly, in contrast to trovafloxacin, two other structurally related quinolone antibiotics ciprofloxacin and levofloxacin did not block PANX1-dependent dye uptake (Fig. 1c), despite all three being equally effective in inhibiting bacterial growth (Extended Data Fig. 1j). Comparing the molecular structure of trovafloxacin with ciprofloxacin and levofloxacin suggested that certain features of trovafloxacin, such as the fluorinated ring at position N1 (Fig. 4h), might contribute to its effect on PANX1 (Figs 1c and 2d). We next asked whether other quinolones with a fluorinated ring at position N1 can inhibit PANX1, and tested difloxacin and tosufloxacin (Fig. 4h) that are currently available commercially in some Asian countries. Although less potent than trovafloxacin, both difloxacin and tosufloxacin partially inhibited ATP release and TO-PRO-3 uptake by apoptotic cells, and promoted the formation of apoptotic bodies (Fig. 4i). These data do not exclude other molecular features of trovafloxacin that may also contribute to PANX1 inhibition, such as the C7 position of the quinolone (Fig. 4h).

Discussion

The data presented here suggest an unexpected but intriguing link between the antibiotic trovafloxacin, eukaryotic pannexin channels,

and a specific step in progression through apoptosis. Specifically, they provide three key insights. First, trovafloxacin, originally marketed by Pfizer, showed tremendous promise, but was linked to severe toxicity and unexplained deaths among children in a trial in Nigeria, and was discontinued[16,35]. Because billions of cells are turned over daily via caspase-mediated apoptosis in the human body as part of normal life, and trovafloxacin can directly inhibit the mammalian PANX1 channels at doses normally achieved in humans, this could, in part, provide a mechanism for trovafloxacin toxicity. Second, our data identify a new role for pannexin channels in maintaining cellular integrity during apoptosis. Fragmentation of apoptotic cells into smaller apoptotic bodies has been seen *in vitro* and *in vivo*[22], but the mechanism(s) regulating this disassembly process is not well understood. The trovafloxacin-mediated blockage of PANX1 uncovered a previously unappreciated step during apoptosis, and revealed an essential and non-redundant role for pannexin channels in the ordered disassembly of apoptotic cells. Third, the dearth of new antibiotics in the pipeline despite the steady increase in drug-resistant bacteria worldwide[33,34] is a significant health threat[36]. Interestingly, quinolone antibiotics such as ciprofloxacin and levofloxacin do not share the inhibitory effect of trovafloxacin towards the mammalian PANX1, and certain features in trovafloxacin might in part explain its toxicity. Further studies on the interaction of quinolones with PANX1 and the associated apoptosis pathways may provide clues to the underlying reasons for the idiosyncratic and irreversible liver toxicity observed with trovafloxacin. Unlocking such clues could re-invigorate industrial interest in the quinolone class for the production of safer, and more potent antibacterials.

METHODS SUMMARY

For drug screening targeting pannexin channels, Jurkat cells were induced to undergo apoptosis (4 h) by anti-Fas (250 ng ml^{-1}) in the presence of 10 μM of each compound from the $LOPAC^{1280}$ library (Sigma-Aldrich). Cells were then stained with TO-PRO-3 and analysed by flow cytometry. Chimeric mice were generated using *Panx1*-targeted embryonic stem cells (from KOMP repository) and crossed to β-actin/Flp mice (JaxMice) to delete the neomycin cassette, and crossed with *Ella*-Cre mice (JaxMice) or *Lck*-Cre mice (JaxMice) to delete PANX1 globally or specifically in thymocytes, respectively. For determining apoptotic cell fragmentation by flow cytometry, Jurkat cells were stained with annexin V-FITC, 7-AAD and TO-PRO-3, whereas thymocytes from *in vivo* experiments were stained similarly with annexin V-FITC, 7-AAD, TO-PRO-3, CD8a conjugated with phycoerythrin and CD4 conjugated with phycoerythrin-Cy7, with gating as per Extended Data Figs 2, 5 and 6. For time-lapse microscopy, cells seeded onto Lab-TekII Chambered Coverglass (Nunc) were imaged live using a Zeiss microscope in a 37 °C/5% CO_2 atmosphere. Whole-cell and inside-out patch recordings were made as described[13,14] (see Methods). A ramp voltage was performed using pCLAMP software and a Digidata 1322A digitizer (Molecular Devices), with data analysed by pCLAMP software.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 December 2013; accepted 12 February 2014.
Published online 12 March 2014.

1. Penuela, S., Gehi, R. & Laird, D. W. The biochemistry and function of pannexin channels. *Biochim. Biophys. Acta* **1828,** 15–22 (2013).
2. Sosinsky, G. E. *et al.* Pannexin channels are not gap junction hemichannels. *Channels (Austin)* **5,** 193–197 (2011).
3. Chen, Y. *et al.* Purinergic signaling: a fundamental mechanism in neutrophil activation. *Sci. Signal.* **3,** ra45 (2010).
4. Seminario-Vidal, L. *et al.* Rho signaling regulates pannexin 1-mediated ATP release from airway epithelia. *J. Biol. Chem.* **286,** 26277–26286 (2011).
5. Séror, C. *et al.* Extracellular ATP acts on P2Y2 purinergic receptors to facilitate HIV-1 infection. *J. Exp. Med.* **208,** 1823–1834 (2011).
6. Billaud, M., Sandilos, J. K. & Isakson, B. E. Pannexin 1 in the regulation of vascular tone. *Trends Cardiovasc. Med.* **22,** 68–72 (2012).
7. Karatas, H. *et al.* Spreading depression triggers headache by activating neuronal Panx1 channels. *Science* **339,** 1092–1095 (2013).
8. Kim, J. E. & Kang, T. C. The P2X7 receptor-pannexin-1 complex decreases muscarinic acetylcholine receptor-mediated seizure susceptibility in mice. *J. Clin. Invest.* **121,** 2037–2047 (2011).
9. MacVicar, B. A. & Thompson, R. J. Non-junction functions of pannexin-1 channels. *Trends Neurosci.* **33,** 93–102 (2010).
10. Lazarowski, E. R. Vesicular and conductive mechanisms of nucleotide release. *Purinergic Signal.* **8,** 359–373 (2012).
11. Dubyak, G. R. Ion homeostasis, channels, and transporters: an update on cellular mechanisms. *Adv. Physiol. Educ.* **28,** 143–154 (2004).
12. Elliott, M. R. *et al.* Nucleotides released by apoptotic cells act as a find-me signal to promote phagocytic clearance. *Nature* **461,** 282–286 (2009).
13. Chekeni, F. B. *et al.* Pannexin 1 channels mediate 'find-me' signal release and membrane permeability during apoptosis. *Nature* **467,** 863–867 (2010).
14. Sandilos, J. K. *et al.* Pannexin 1, an ATP release channel, is activated by caspase cleavage of its pore-associated C-terminal autoinhibitory region. *J. Biol. Chem.* **287,** 11303–11311 (2012).
15. Qu, Y. *et al.* Pannexin-1 is required for ATP release during apoptosis but not for inflammasome activation. *J. Immunol.* **186,** 6553–6561 (2011).
16. Liu, H. H. Safety profile of the fluoroquinolones: focus on levofloxacin. *Drug Saf.* **33,** 353–369 (2010).
17. Stahlmann, R. & Lode, H. Safety considerations of fluoroquinolones in the elderly: an update. *Drugs Aging* **27,** 193–209 (2010).
18. Bruzzone, R., Barbe, M. T., Jakob, N. J. & Monyer, H. Pharmacological properties of homomeric and heteromeric pannexin hemichannels expressed in *Xenopus* oocytes. *J. Neurochem.* **92,** 1033–1043 (2005).
19. Ma, W., Hui, H., Pelegrin, P. & Surprenant, A. Pharmacological characterization of pannexin-1 currents expressed in mammalian cells. *J. Pharmacol. Exp. Ther.* **328,** 409–418 (2009).
20. Vincent, J. *et al.* Pharmacokinetics and safety of trovafloxacin in healthy male volunteers following administration of single intravenous doses of the prodrug, alatrofloxacin. *J. Antimicrob. Chemother.* **39** (suppl. B), 75–80 (1997).
21. Teng, R., Liston, T. E. & Harris, S. C. Multiple-dose pharmacokinetics and safety of trovafloxacin in healthy volunteers. *J. Antimicrob. Chemother.* **37,** 955–963 (1996).
22. Wickman, G., Julian, L. & Olson, M. F. How apoptotic cells aid in the removal of their own cold dead bodies. *Cell Death Differ.* **19,** 735–742 (2012).
23. Moss, D. K., Betin, V. M., Malesinski, S. D. & Lane, J. D. A novel role for microtubules in apoptotic chromatin dynamics and cellular fragmentation. *J. Cell Sci.* **119,** 2362–2374 (2006).
24. Nagata, S., Hanayama, R. & Kawane, K. Autoimmunity and the clearance of dead cells. *Cell* **140,** 619–630 (2010).
25. Bortner, C. D. & Cidlowski, J. A. Caspase independent/dependent regulation of K^+, cell shrinkage, and mitochondrial membrane potential during lymphocyte apoptosis. *J. Biol. Chem.* **274,** 21953–21962 (1999).
26. Vanden Berghe, T. *et al.* Determination of apoptotic and necrotic cell death *in vitro* and *in vivo*. *Methods* **61,** 117–129 (2013).
27. Silverman, W., Locovei, S. & Dahl, G. Probenecid, a gout remedy, inhibits pannexin 1 channels. *Am. J. Physiol. Cell Physiol.* **295,** C761–C767 (2008).
28. Cohen, J. J., Duke, R. C., Fadok, V. A. & Sellins, K. S. Apoptosis and programmed cell death in immunity. *Annu. Rev. Immunol.* **10,** 267–293 (1992).
29. Baroja-Mazo, A., Barbera-Cremades, M. & Pelegrin, P. The participation of plasma membrane hemichannels to purinergic signaling. *Biochim. Biophys. Acta* **1828,** 79–93 (2013).
30. Coleman, M. L. *et al.* Membrane blebbing during apoptosis results from caspase-mediated activation of ROCK I. *Nature Cell Biol.* **3,** 339–345 (2001).
31. Sebbagh, M. *et al.* Caspase-3-mediated cleavage of ROCK I induces MLC phosphorylation and apoptotic membrane blebbing. *Nature Cell Biol.* **3,** 346–352 (2001).
32. Andriole, V. T. The quinolones: past, present, and future. *Clin. Infect. Dis.* **41** (suppl. 2), S113–S119 (2005).
33. Centers for Disease Control and Prevention. Antibiotic resistance threats in the United States, 2013. http://www.cdc.gov/drugresistance/threat-report-2013/ (US Department of Health and Human Services, 2013).
34. Department of Health. UK five year antimicrobial resistance strategy 2013 to 2018. https://www.gov.uk/government/publications/uk-5-year-antimicrobial-resistance-strategy-2013-to-2018 (2013).
35. Ahmad, K. Drug company sued over research trial in Nigeria. *Lancet* **358,** 815 (2001).
36. Spellberg, B. *et al.* The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. *Clin. Infect. Dis.* **46,** 155–164 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank B. Isakson, M. Billaud, J. Lannigan and other colleagues for discussions. Grants from the US National Institutes of Health (NIGMS 107848 to K.S.R. and D.B.) and the National Health & Medical Research Council of Australia (I.K.H.P.), a training fellowship to I.K.H.P., and the American Heart Association (J.M.K.) supported this work.

Author Contributions I.K.H.P. designed, performed and analysed most of the experiments with input from K.S.R. Y.-H.C. performed and analysed the patch-clamp studies with input from D.A.B. A.J.A. generated the PANX1 knockout mice and Jurkat cell lines expressing PANX1. J.M.K. helped with the drug screen and microscopy experiments. I.J.J. performed primary T-cell isolation and quantitative PCR. A.J.A. and I.J.J. assisted with the initial characterization of PANX1-deficient mice. I.K.H.P. and K.S.R. wrote the manuscript with input from co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.S.R. (Ravi@virginia.edu).

METHODS

Reagents. Library of Pharmacologically Active Compounds (LOPAC[1280]; http://www.sigmaaldrich.com/chemistry/drug-discovery/validation-libraries/lopac1280-navigator.html), trovafloxacin, ciprofloxacin, levofloxacin, difloxacin, tosufloxacin, carbenoxolone, probenecid, Y-27632, blebbistatin, cytochalasin D, purified nucleotides, suramin and dexamethasone were obtained from Sigma-Aldrich. 7-AAD and TO-PRO-3 were purchased from Invitrogen. Annexin V-FITC, CD8a-PE (clone 53-6.7), CD4-PE-Cy7 (clone RM4-5) and anti-mouse CD16/CD32 (clone 93) were obtained from eBioscience. Other reagents were obtained as follows: anti-Fas (clone CH11, Millipore), z-VAD-FMK (Enzo Life Sciences), Q-VD-OPH (SM Biochemicals) and recombinant apyrase (New England Biolabs).

Induction of apoptosis. Jurkat cells in RPMI with 1% BSA were treated with 250 ng ml^{-1} anti-Fas (clone CH11) or 150 mJ cm^{-2} ultraviolet C irradiation (Stratalinker). Primary thymocytes (collected from 5- to 7-week-old C57BL/6 mice, male and female) were treated with 50 μM dexamethasone for 5 h. LR73 cells were treated with 150 mJ cm^{-2} ultraviolet C irradiation for 6 h. All treatments were incubated for indicated times at 37 °C, 5% CO_2, unless noted otherwise.

Drug screening for regulators of PANX1 function. Jurkat cells (10^6 cells ml^{-1} in DMEM + 0.5% BSA) were induced to undergo apoptosis by anti-Fas treatment (250 ng ml^{-1}) in the presence of 10 μM of compounds in the LOPAC[1280] (Sigma-Aldrich) for 4 h at 37 °C, 5% CO_2. Cells were then stained with TO-PRO-3 (0.67 μM) for 10 min at room temperature and immediately placed on ice before analysis on a BD FACSCanto flow cytometer. The resultant flow cytometric data were analysed by FlowJo software (Tree Star).

Determining cell viability by flow cytometry and ImageStream. Samples were stained with annexin V-FITC, 7-AAD and TO-PRO-3 in annexin V binding buffer for 10 min at room temperature and immediately placed on ice before analysis on a BD FACSCanto flow cytometer, with the resultant flow cytometry data analysed by FlowJo software. Samples were also prepared the same way before analysis on an ImageStreamX Mark II (Amnis), with the resultant data analysed by IDEAS software (Amnis).

For TO-PRO-3 uptake assay following acute drug treatment, Jurkat cells in RPMI/1% BSA were treated with 500 ng ml^{-1} anti-Fas (clone CH11) for 4 h at 37 °C, 5% CO_2. Following induction of apoptosis, cells were treated with pan caspase inhibitor Q-VD-OPH (50 μM) and the indicated concentrations of compounds for 20 min at 37 °C, 5% CO_2. Cells were subsequently stained with annexin V-FITC, 7-AAD and TO-PRO-3 and analysed on a FACSCanto flow cytometer. The resultant flow cytometry data were analysed by FlowJo software.

Generation of pannexin 1 knockout mice. *Panx1*-targeted embryonic stem cells (EPD0309-3-B01) were obtained from the Knockout Mouse Project (KOMP) Repository. After blastocyst injections (performed by the transgenic core facility at the University of Virginia), chimaeras were bred with C57BL/6J mice (JaxMice). The resultant offspring were crossed with β-actin/Flp mice (JaxMice) to delete the neomycin cassette, thereby generating mice carrying a floxed *Panx1* exon 3 allele (*Panx1$^{fl/fl}$*) (Extended Data Fig. 7). To generate *Panx1* global knockout mice, *Panx1$^{fl/fl}$* mice were crossed with *Ella*-Cre mice (JaxMice) expressing Cre from the two-cell stage of embryonic development to delete *Panx1* exon 3 in all tissues[37]. Resulting mice were crossed with C57BL/6J mice to remove Flp and Cre from the background. To generate mice that carry deletion of *Panx1* specifically in thymocytes, *Panx1$^{fl/fl}$* mice were crossed with *Lck*-Cre mice (JaxMice) expressing Cre under the *Lck* proximal promoter, which mediates deletion of *Panx1* exon 3 from the double-negative stage of thymocyte development[38].

Microscopy. Live imaging was performed on a Zeiss microscope using a ×40 or ×60 oil immersion objective in a 37 °C/5% CO_2 atmosphere. Jurkat cells at 5×10^5 cells ml^{-1} in RPMI+1% BSA (600 μl) were seeded onto Lab-Tek II Chambered Coverglass (Nunc) by two successive centrifugations at 20*g* for 1 min before imaging. LR73 fibroblasts were cultured overnight on Chambered Coverglass before imaging. In certain experiments, TO-PRO-3 and annexin V-FITC were also present during the imaging process.

Bacterial growth assay. Overnight *Escherichia coli* culture was diluted 100-fold in LB medium before incubation with quinolones for 6 h. Bacterial growth was monitored by the absorbance at 600 nm using a FlexStation 3 plate reader, with the resultant data analysed using SoftMaxPro 5.4 software.

Nucleotide measurement. ATP in apoptotic cell supernatants was measured using a luciferase/luciferin assay (CellTiter-Glo; Promega) according to manufacturer's instructions.

Caspase activation. Caspase 3/7 activity assays were performed with the Caspase-Glo 3/7 (Promega) reagents, in accordance with the manufacturer's instructions.

Immunoblotting. Samples were analysed by SDS–PAGE and immunoblotted using the following dilutions: anti-GFP (1:1,000; Santa Cruz), anti-ERK2 (1:3,000; Santa Cruz), anti-Caspase 3 (1:2,500; Santa Cruz) and affinity purified rabbit anti-mouse PANX1 (0.2 μg ml^{-1}).

Patch-clamp analysis. Electrophysiological recordings were made at room temperature using an Axopatch 200B amplifier (Molecular Devices). Whole-cell recordings were carried out in Jurkat cells or transiently transfected HEK293T cells with borosilicate glass patch pipettes (3–5 MΩ). Bath solution contained 140 mM NaCl, 3 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 10 mM HEPES, and 10 mM glucose (pH 7.3). Pipette solution was composed of 30 mM tetraethylammonium chloride, 100 mM $CsMeSO_4$, 4 mM NaCl, 1 mM $MgCl_2$, 0.5 mM $CaCl_2$, 10 mM HEPES, 10 mM EGTA, 3 mM ATP-Mg, and 0.3 mM GTP-Tris (pH 7.3). A ramp voltage command was applied using pCLAMP software and a Digidata 1322A digitizer (Molecular Devices). Data were analysed using pCLAMP software. Results were presented as means ± s.e.m. The dose–response curve was fitted and IC_{50} was determined using Prism 5 software. Inside-out patch recordings were obtained using patch pipettes (2–4 MΩ) filled with the bath solution (described above) in a perfusate containing 150 mM CsCl, 5 mM EGTA, 1 mM MgCl2, 10 mM HEPES (pH 7.3). Currents were evoked from a holding potential of −50 mV. All-point histograms and NPo analysis was performed using pCLAMP software. Connexin 43 currents were measured in divalent-free bath solution[13,14,39].

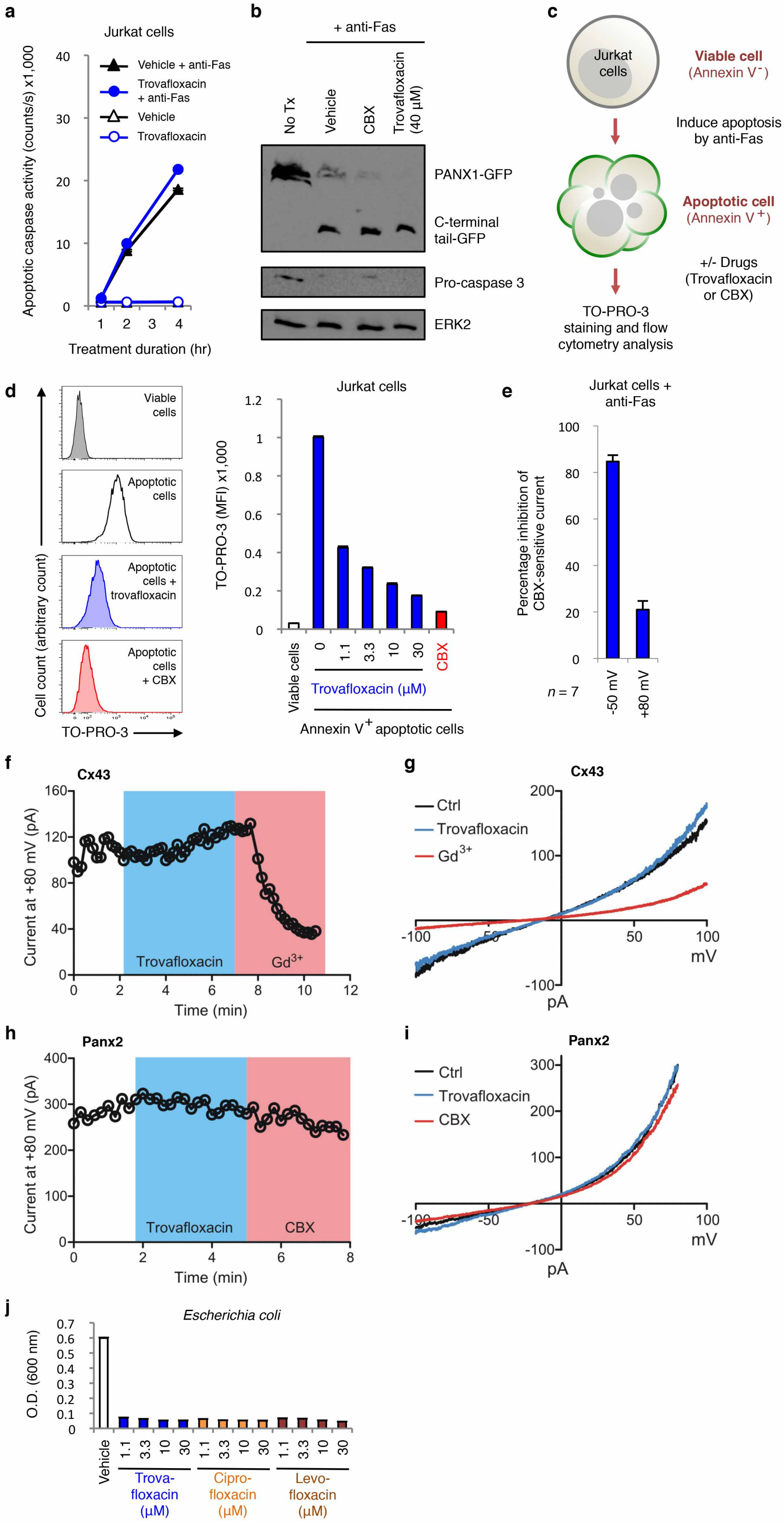
DNA fragmentation assay. DNA laddering during apoptosis was characterized by agarose gel electrophoresis[40]. In brief, 60,000 Jurkat cells were induced to undergo apoptosis by anti-Fas treatment for 2 h and lysed in TES lysis buffer containing RNase for 2 h at 37 °C. Cell lysates were subsequently treated with proteinase K for 18 h at 50 °C and DNA fragmentation was visualized by agarose gel electrophoresis.

Quantitative PCR. cDNA was synthesized from 50 ng of RNA isolated from primary thymocytes (RNeasy, Qiagen) using Superscript III (Invitrogen). Quantitative PCR was performed on the ABI StepOnePlus Real-time PCR instrument with TaqMan probes (Applied Biosystems). Levels of mouse *Panx1* mRNA are normalized to *Gapdh*. TaqMan probes used were: mouse *Panx1* (Mm00450900_m1) and mouse *Gapdh* (4352339E-1207039).

***In vivo* model of apoptosis.** 5-week-old C57BL/6 mice (male and female) were injected intraperitoneally with 12.5 mg kg^{-1} dexamethasone. Trovafloxacin (60 mg kg^{-1}; intraperitoneally) was administered 10 min prior to dex injection (12.5 mg kg^{-1}; intraperitoneally), and trovafloxacin was also given at 2 and 4 h and the analysis performed at 6 h. Thymocytes were collected 6 h post dexamethasone injection, stained with annexin V-FITC, 7-AAD, TO-PRO-3, CD8a-PE and CD4-PE-Cy7, and analysed on a BD FACSCanto flow cytometer. The University of Virginia Animal Care and Use Committee approved all animal experiments.

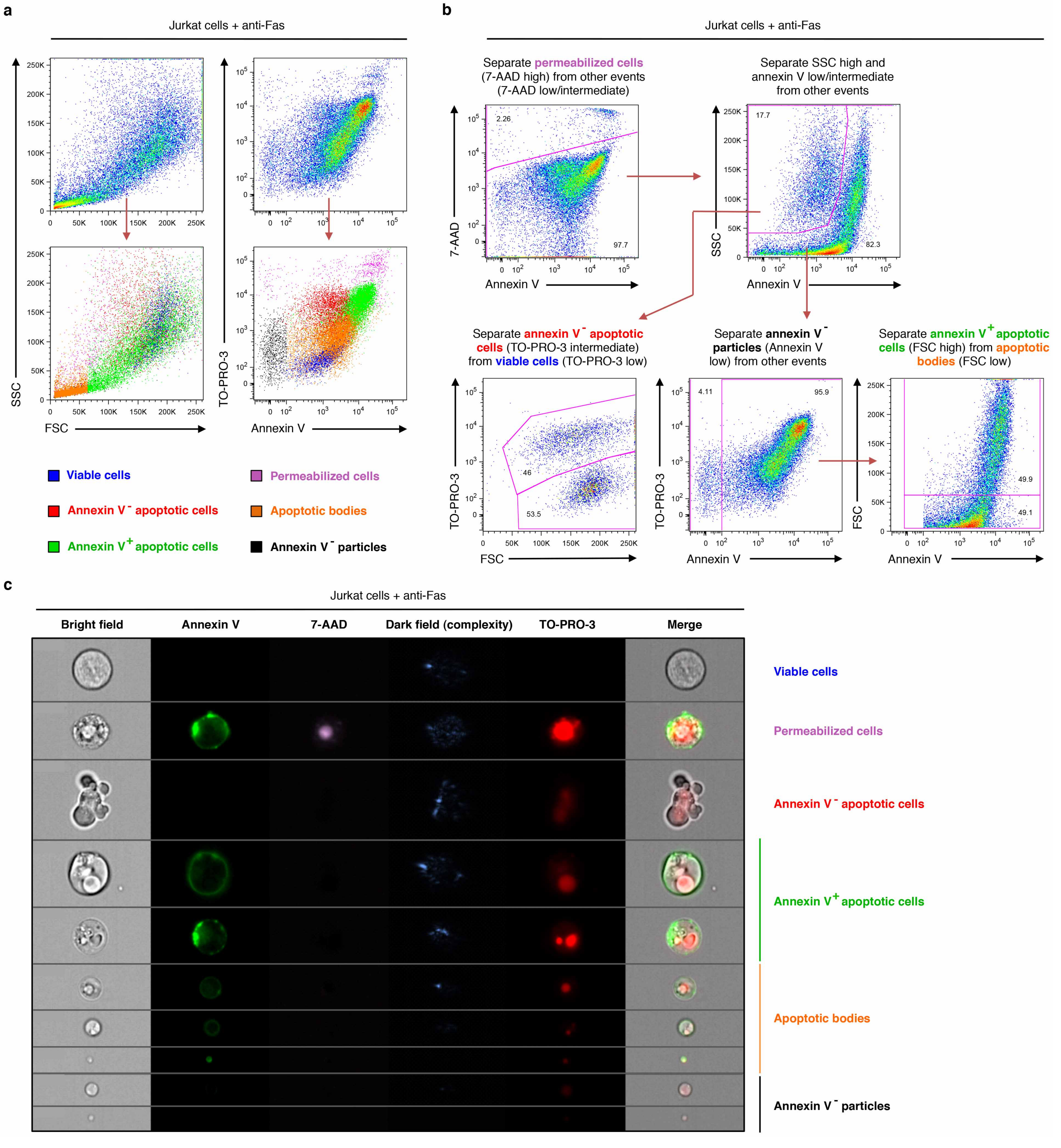
Statistical analyses. Data are presented as means ± s.e.m. Statistical significance for comparisons was determined by unpaired Student's two-tailed *t*-test. A *P* value less than 0.05 was considered statistically significant. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

37. Lakso, M. *et al.* Efficient *in vivo* manipulation of mouse genomic sequences at the zygote stage. *Proc. Natl Acad. Sci. USA* **93,** 5860–5865 (1996).
38. Lee, P. P. *et al.* A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. *Immunity* **15,** 763–774 (2001).
39. Harris, A. L. Connexin channel permeability to cytoplasmic molecules. *Prog. Biophys. Mol. Biol.* **94,** 120–143 (2007).
40. Henry, C. M., Hollville, E. & Martin, S. J. Measuring apoptosis by microscopy and flow cytometry. *Methods* **61,** 90–97 (2013).

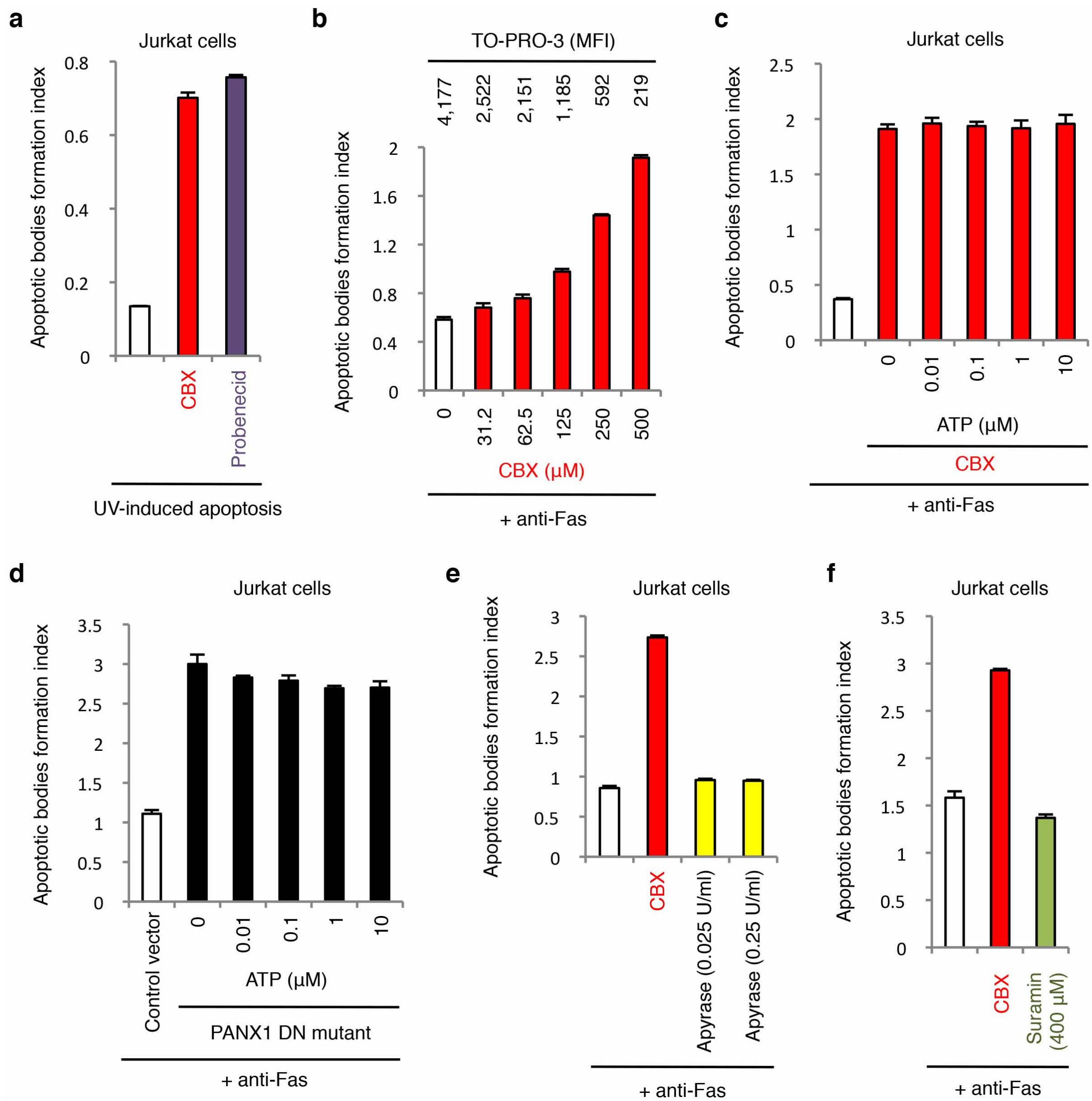


a
Jurkat cells
Vehicle + anti-Fas
Trovafloxacin + anti-Fas
Vehicle
Trovafloxacin
Apoptotic caspase activity (counts/s) x1,000
Treatment duration (hr)
b
+ anti-Fas
No Tx
Vehicle
CBX
Trovafloxacin (40 μM)
PANX1-GFP
C-terminal tail-GFP
Pro-caspase 3
ERK2
c
Jurkat cells
Viable cell (Annexin V⁻)
Induce apoptosis by anti-Fas
Apoptotic cell (Annexin V⁺)
+/- Drugs (Trovafloxacin or CBX)
TO-PRO-3 staining and flow cytometry analysis
d
Viable cells
Apoptotic cells
Apoptotic cells + trovafloxacin
Apoptotic cells + CBX
Cell count (arbitrary count)
TO-PRO-3
Jurkat cells
TO-PRO-3 (MFI) x1,000
Viable cells
Trovafloxacin (μM)
CBX
Annexin V⁺ apoptotic cells
e
Jurkat cells + anti-Fas
Percentage inhibition of CBX-sensitive current
n = 7
-50 mV
+80 mV
f
Cx43
Current at +80 mV (pA)
Trovafloxacin
Gd³⁺
Time (min)
g
Cx43
Ctrl
Trovafloxacin
Gd³⁺
mV
pA
h
Panx2
Current at +80 mV (pA)
Trovafloxacin
CBX
Time (min)
i
Panx2
Ctrl
Trovafloxacin
CBX
mV
pA
j
Escherichia coli
O.D. (600 nm)
Vehicle
Trova-floxacin (μM)
Cipro-floxacin (μM)
Levo-floxacin (μM)

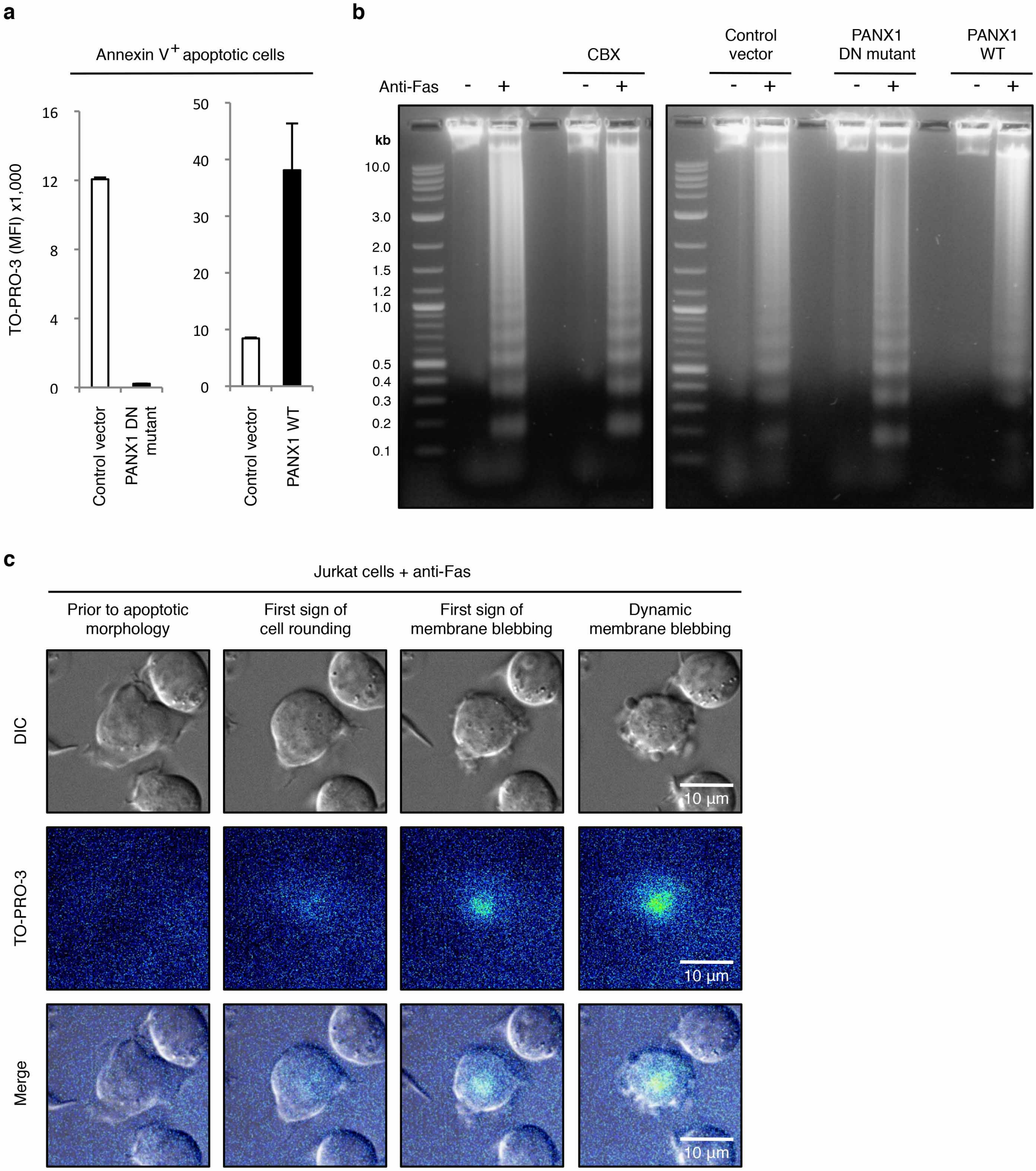
Extended Data Figure 1 | Trovafloxacin does not block caspase activation or inhibit connexin 43 (Cx43) or pannexin 2 (Panx2) membrane currents. **a**, Caspase 3/7 activation in Jurkat cells undergoing apoptosis is not altered by treatment with trovafloxacin (40 μM) ($n = 3$). **b**, Proteolytic cleavage of PANX1–GFP during apoptosis is not inhibited by trovafloxacin (40 μM) or CBX (500 μM) treatment. **c**, Schematic diagram for the acute treatment of apoptotic cells with trovafloxacin or CBX. **d**, Acute trovafloxacin treatment inhibits TO-PRO-3 uptake by apoptotic Jurkat cells. Left, histograms showing TO-PRO-3 uptake by viable cells, apoptotic cells, or apoptotic cells treated with trovafloxacin or CBX (500 μM) post induction of apoptosis and analysed by flow cytometry. Right, uptake of TO-PRO-3 presented as (median fluorescence intensity, MFI) of viable cells or apoptotic cells ($n = 3$). **e**, Inhibition of CBX-sensitive current in apoptotic cells treated with trovafloxacin (20 μM), as measured by whole-cell patch-clamp recording ($n = 7$). **f**, Patch-clamp recordings from HEK293T cells expressing Cx43 and receiving indicated treatments. Whole-cell current at +80 mV is shown under conditions when bath solution was perfused with trovafloxacin (20 μM, blue shading) or gadolinium (Gd^{3+}) (100 μM, pink shading). **g**, Current-voltage relationships of Cx43 current in HEK293T cells treated with or without trovafloxacin (20 μM) or Gd^{3+} (100 μM), with the current measured over a range of voltages. Exemplar traces in **f** and **g** are representative of 14 cells per group. **h**, Patch-clamp recordings from HEK293T cells expressing mouse Panx2 and receiving indicated treatments. Whole-cell current at +80 mV is shown under conditions when bath solution was perfused with trovafloxacin (20 μM, blue shading) or carbenoxolone (CBX) (50 μM, pink shading). **i**, Current-voltage relationships of Panx2 current in HEK293T cells treat with or without trovafloxacin (20 μM) or CBX (50 μM), with the current measured over a range of voltages. Exemplar traces in **h** and **i** are representative of 4 cells per group. **j**, Trovafloxacin, ciprofloxacin and levofloxacin inhibit bacterial growth. *Escherichia coli* growth (as measured by absorbance at 600 nm) in the presence of indicated concentrations of quinolones ($n = 3$). Error bars represent s.e.m.



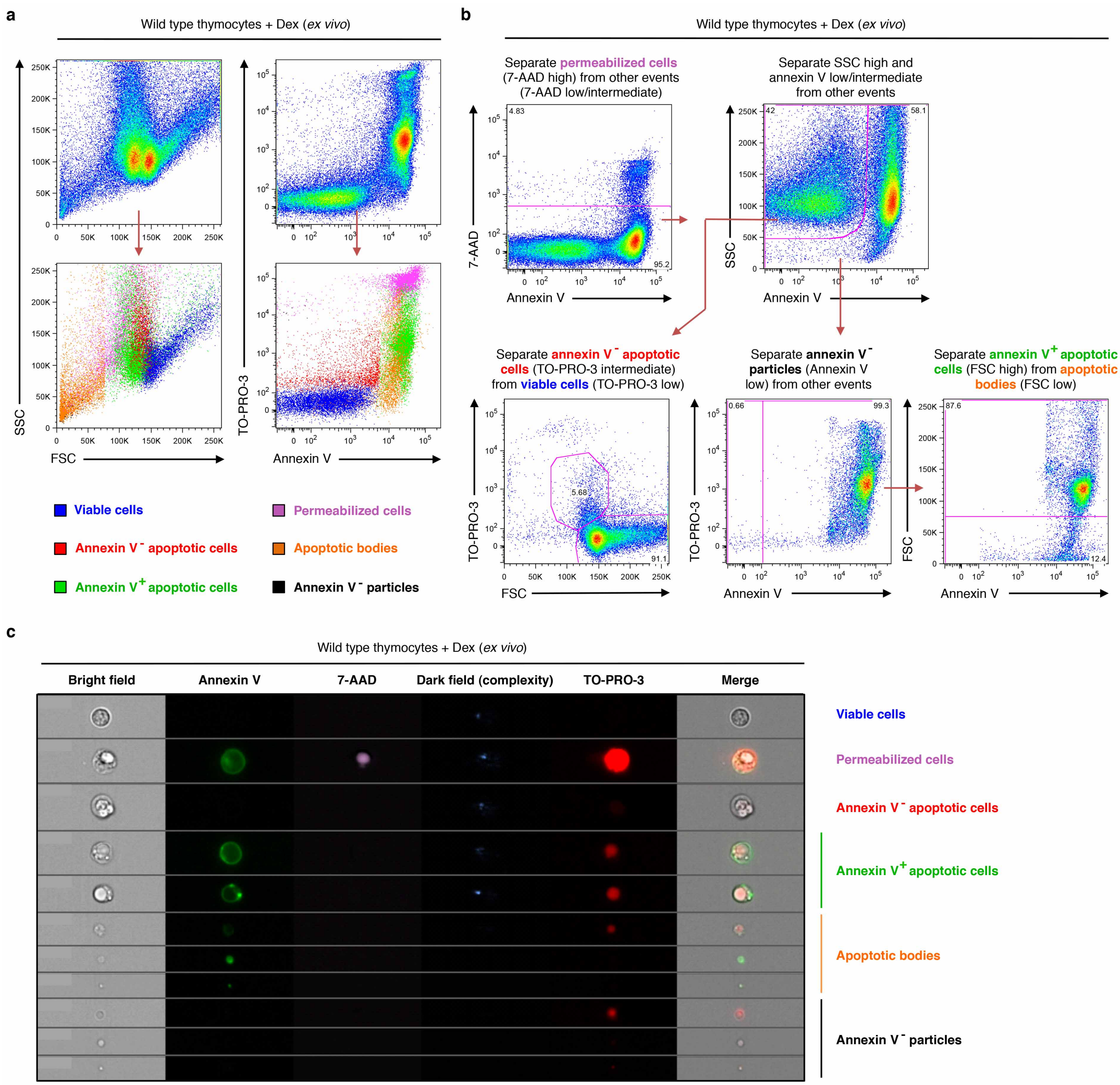
Extended Data Figure 2 | Electronic gating strategy for the separation of different cellular and subcellular population of Jurkat cells undergoing apoptosis *in vitro*. **a**, Flow cytometric analysis showing each type of particles gated (see **b** below) has a distinctive level of cellular complexity (side scatter, SSC), cell size (forward scatter, FSC) as well as TO-PRO-3 (indicative of caspase-mediated activation of pannexin 1 channels), 7-AAD (indicative of membrane integrity) and annexin V (indicative of phosphatidylserine exposure) staining. **b**, Flow cytometry gating strategy used to distinguish viable cells, annexin V^- apoptotic cells, annexin V^+ apoptotic cells, annexin V^- particles, and apoptotic bodies. **c**, ImageStream analysis of particles gated using the same strategy as described in **b**. Representative images for each type of particles are shown. Jurkat cells were induced to undergo apoptosis by anti-Fas treatment (2 h) in all indicated experiments.



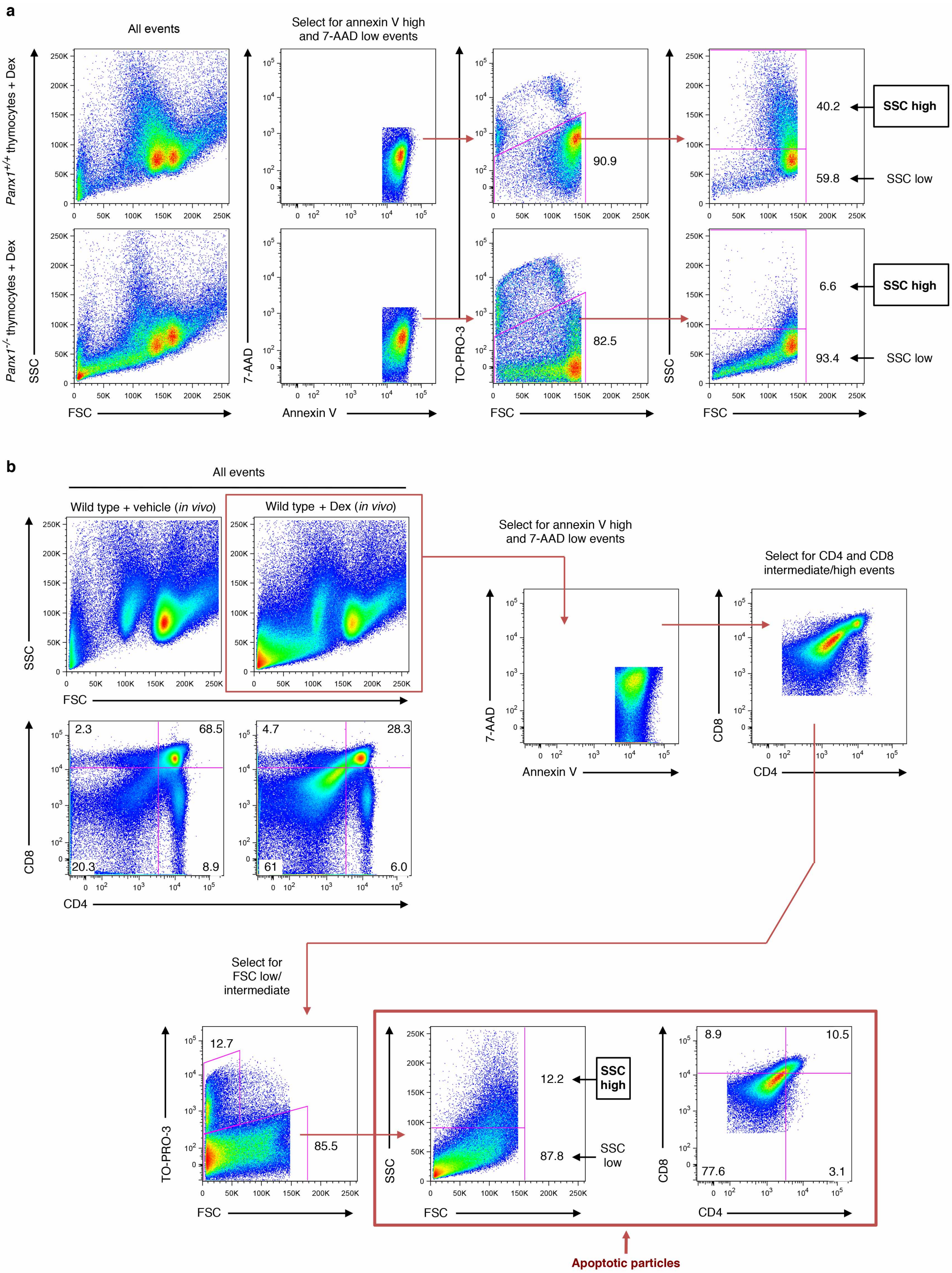
Extended Data Figure 3 | Inhibition of pannexin 1 promotes the formation of apoptotic bodies via a mechanism independent of extracellular ATP. **a**, CBX and probenecid enhance the generation of apoptotic bodies from cells undergoing ultraviolet-induced apoptosis ($n = 3$). **b**, Formation of apoptotic bodies after treatment with the indicated concentrations of CBX ($n = 3$). The corresponding TO-PRO-3 uptake by annexin V^+ apoptotic cells at each CBX concentration is shown above the respective bars. **c**, **d**, Addition of exogenous ATP during apoptosis induction does not inhibit formation of apoptotic bodies in CBX-treated cells ($n = 3$) (**c**) or cells stably expressing the dominant-negative PANX1 mutant (PANX1 DN mutant) ($n = 3$) (**d**). **e**, Removal of extracellular ATP by apyrase does not promote formation of apoptotic bodies ($n = 3$). **f**, P2Y receptor antagonist suramin does not promote formation of apoptotic bodies ($n = 3$). Jurkat cells were induced to undergo apoptosis by anti-Fas treatment in all indicated experiments. Error bars represent s.e.m.



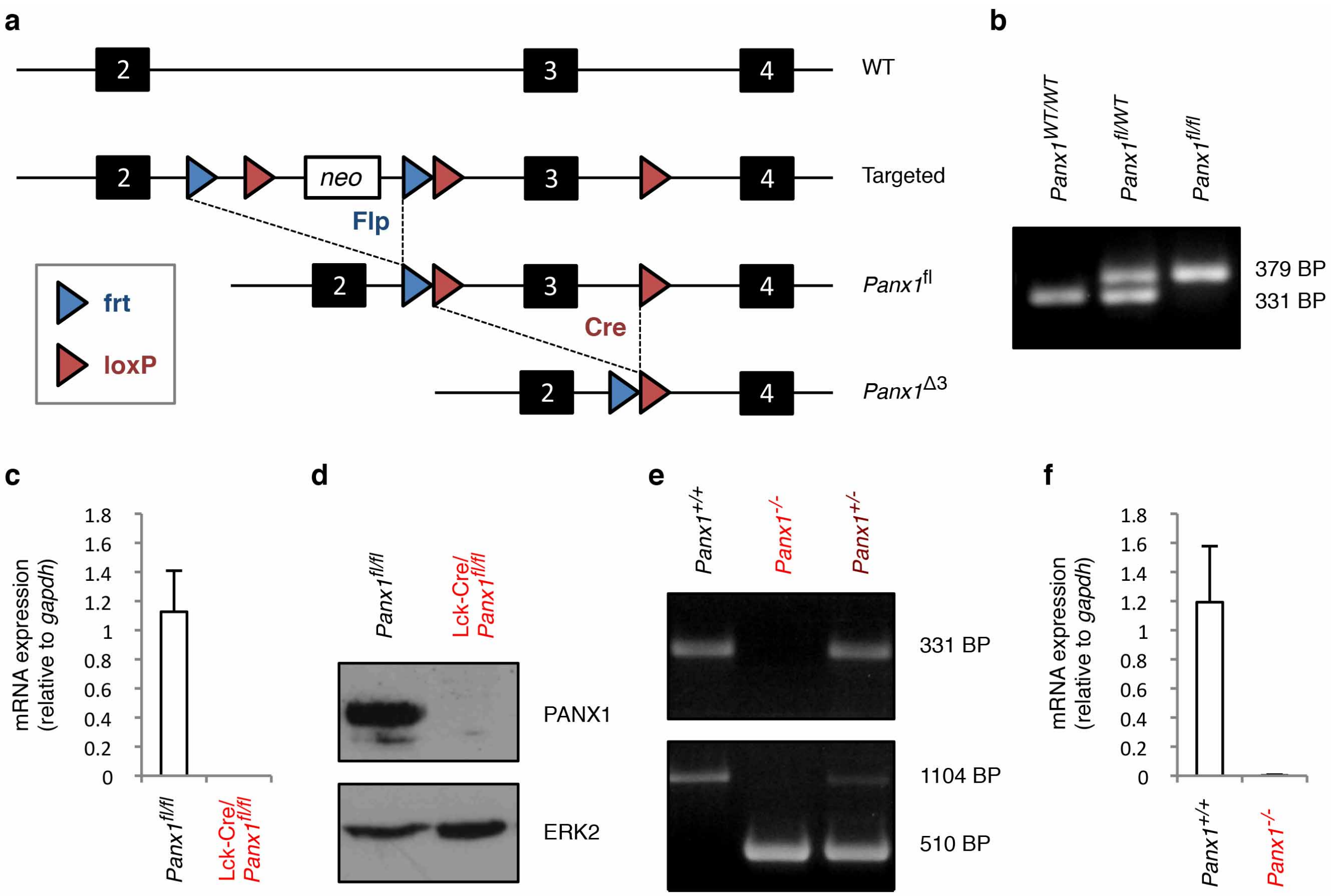
Extended Data Figure 4 | Pannexin 1 activity does not affect DNA fragmentation during apoptosis. **a**, **b**, TO-PRO-3 dye uptake ($n = 3$) (**a**) and DNA fragmentation (**b**) were assessed in Jurkat cells stably expressing the control vector, the dominant-negative PANX1 mutant (PANX1 DN mutant) or wild-type PANX1 (PANX1 WT). DNA fragmentation from cells induced to undergo apoptosis and treated with or without 500 µM CBX is also shown in **b**. **c**, Time-lapse images monitoring TO-PRO-3 dye uptake during progression of apoptosis in Jurkat cells with normal PANX1 function show that TO-PRO-3 uptake occurs before initiation of membrane blebbing. Jurkat cells were induced to undergo apoptosis by anti-Fas treatment (2 h). Error bars in **a** represent s.e.m.



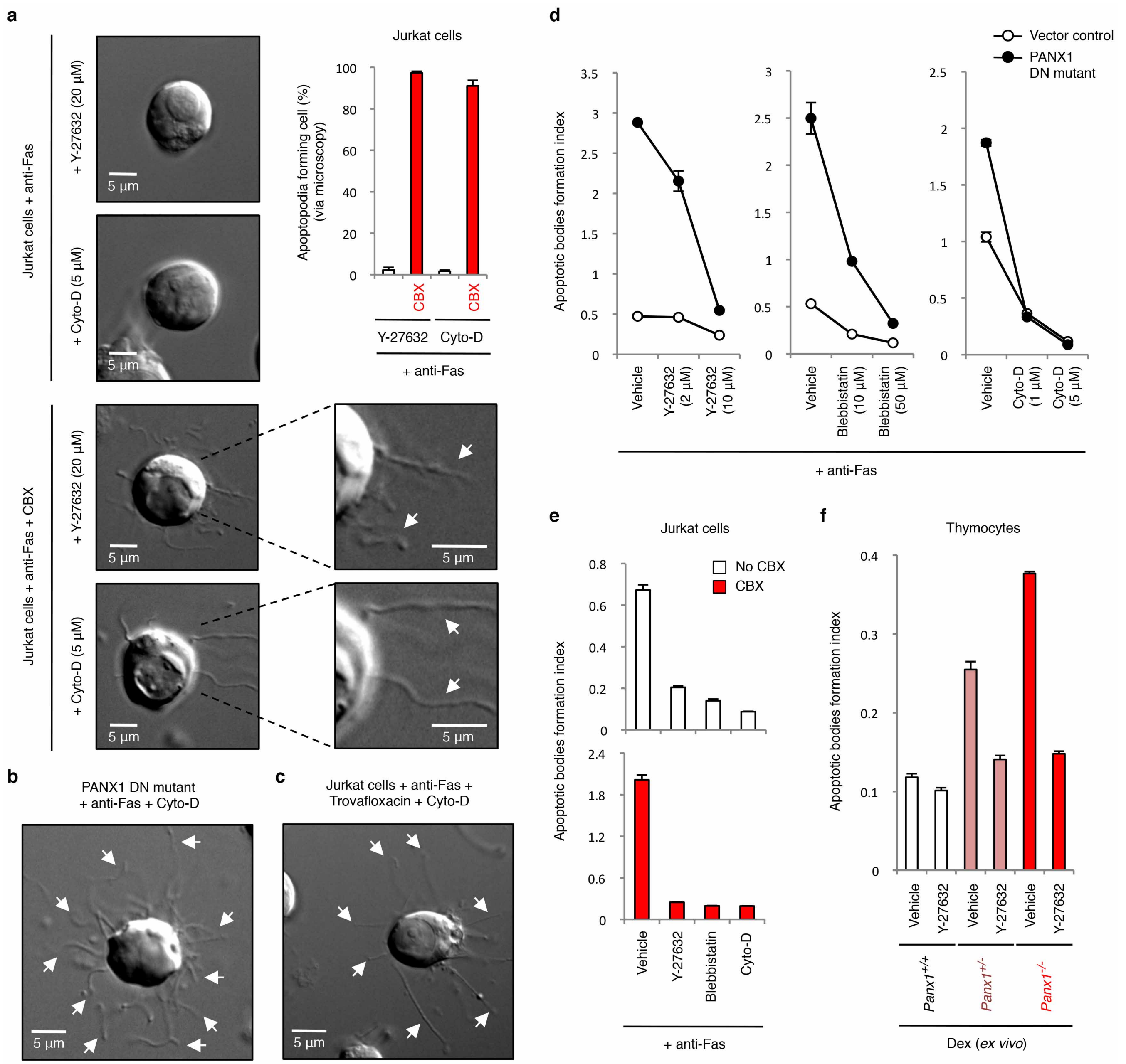
Extended Data Figure 5 | Electronic gating strategy for the separation of different cellular and subcellular populations of primary thymocytes undergoing apoptosis *ex vivo*. **a**, Flow cytometry analysis showing each type of particle gated according to **b** has a distinctive level of SSC, FSC as well as TO-PRO-3 and annexin V staining. **b**, Flow cytometry analysis showing electronic gating strategy used to distinguish viable cells, annexin V^{-} apoptotic cells, annexin V^{+} apoptotic cells, annexin V^{-} particles, and apoptotic bodies. **c**, ImageStream analysis of particles gated using the same strategy as described in **b**. Representative images for each type of particle are shown. Primary mouse thymocytes were induced to undergo apoptosis by dexamethasone (Dex) treatment in all indicated experiments.



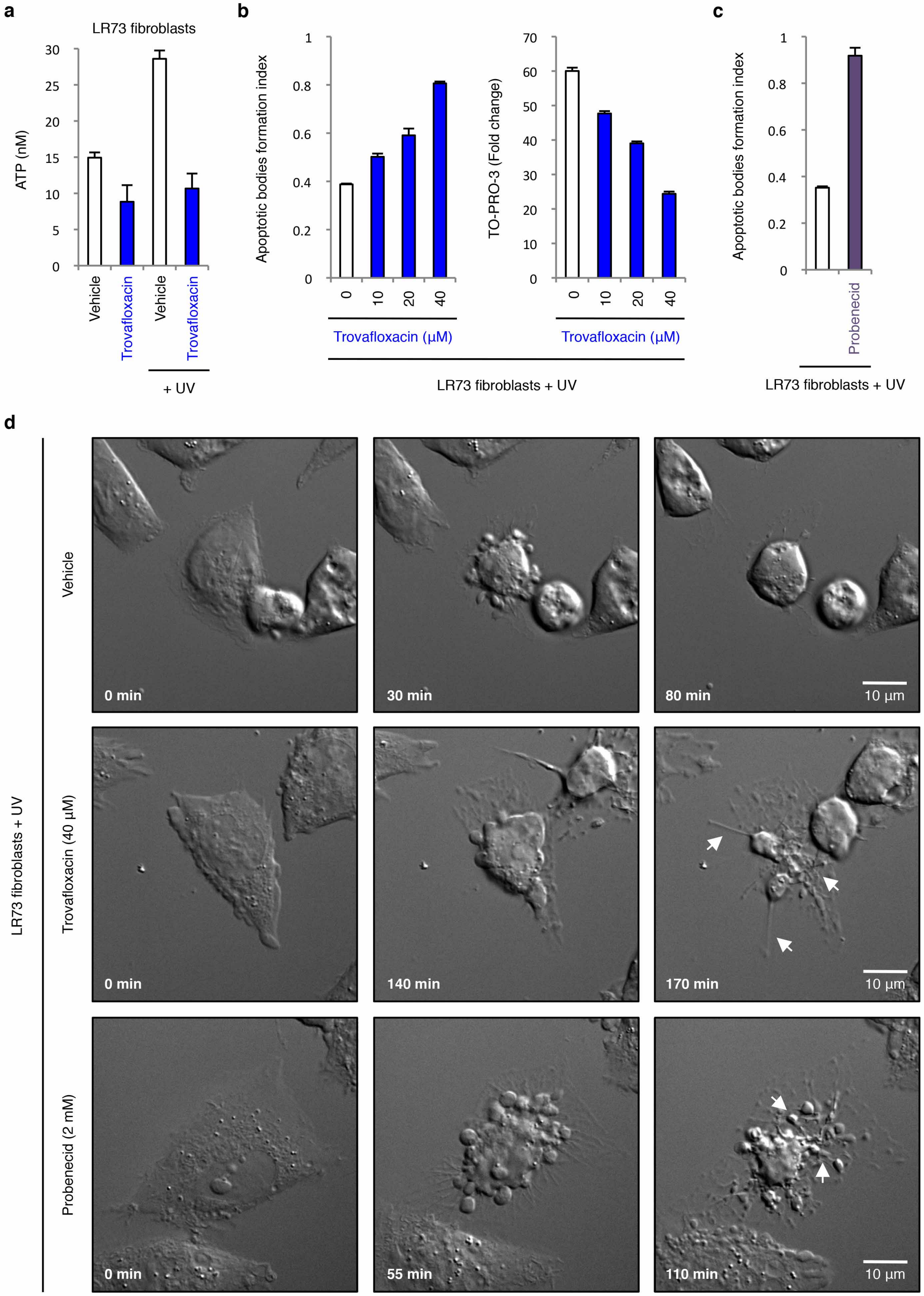
Extended Data Figure 6 | Electronic gating strategy for analysing the complexity of subcellular apoptotic particles generated *ex vivo* and *in vivo*. **a**, Flow cytometry analysis showing electronic gating strategy used to distinguish annexin V^{high}, 7-AAD^{low} subcellular particles generated from primary mouse thymocytes induced to undergo apoptosis via dexamethasone treatment. Subcellular apoptotic particles with high complexity (SSC high) or low complexity (SSC low) are gated as shown. **b**, Flow cytometry analysis showing electronic gating strategy used to distinguish different subsets of apoptotic cell-derived particles generated in the thymus of mice injected intraperitoneally with dexamethasone (6 h). Annexin V^{high}, 7-AAD^{low}, CD4/CD8^{intermediate} particles were initially selected and subsequently gated based on forward scatter (FSC, indicative of cell size). Apoptotic particles of interest (as indicated) are therefore defined as annexin V^{high}, 7-AAD^{low}, CD4/CD8^{intermediate} and FSC$^{low/intermediate}$.



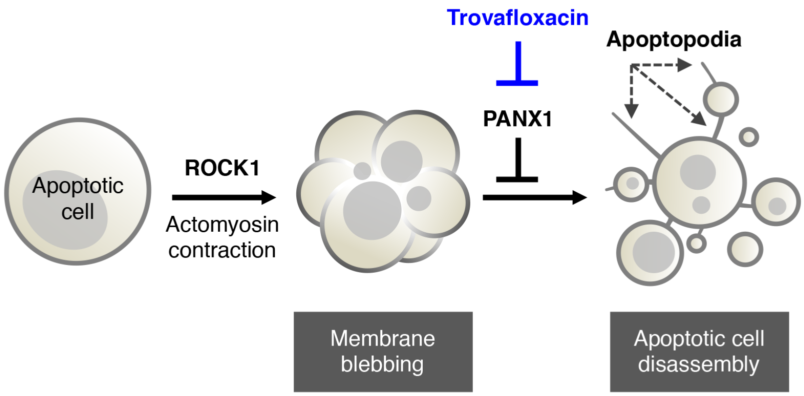
Extended Data Figure 7 | Generation of conditional and global pannexin 1-deficient mice. **a**, Strategy for deletion of neomycin cassette and exon 3 of *Panx1*. **b**, Identification of mice with floxed *Panx1* loci, assessed by PCR. **c**, mRNA levels of *Panx1* in $CD4^+$ thymocytes relative to *Gapdh*. $n = 3$ mice per group. **d**, Immunoblotting of lysates from thymocytes with the indicated genotypes. **e**, Identification of mice with wild-type, heterozygous and homozygous *Panx1*-targeted loci, assessed by PCR. **f**, mRNA levels of *Panx1* in thymocytes with the indicated genotypes relative to *Gapdh*. $n = 3$ mice per group. Error bars in **c** and **f** represent s.e.m.



Extended Data Figure 8 | Formation of apoptotic bodies but not string-like apoptopodia structures is dependent on actomyosin contraction. **a**, Time-lapse images monitoring apoptotic cell morphology of cells treated with or without CBX (500 µM) and in the presence of actomyosin contraction inhibitors. Top right, percentage of apoptotic cells forming string-like apoptopodia structures (387, 414, 459 and 372 apoptotic cells were analysed for Y-27632, Y-27632+CBX, Cyto-D and Cyto-D+CBX-treated cells, respectively, from three independent experiments). **b**, **c**, Time-lapse images monitoring apoptotic cell morphology of cells stably expressing the dominant-negative PANX1 mutant (PANX DN mutant) (**b**) or treated with 40 µM trovafloxacin (**c**) in the presence of Cyto-D (5 µM). **d**, Inhibitors of blebbing, Y-27632, blebbistatin, or cytochalasin D (Cyto-D) reduce the formation of apoptotic bodies in Jurkat cells expressing PANX1 DN mutant ($n = 3$). **e**, Generation of apoptotic bodies by dying cells treated with Y-27632 (10 µM), blebbistatin (50 µM) and Cyto-D (5 µM). Cells were induced to undergo apoptosis in the presence or absence of CBX (500 µM) ($n = 3$). **f**, The enhanced formation of apoptotic bodies in apoptotic thymocytes from mice with PANX1 deficiency is also blunted by the ROCK inhibitor Y-27632 (10 µM) that blocks membrane blebbing ($n = 3$). Error bars represent s.e.m. Scale bars, 5 µm. Arrows, apoptopodia.



Extended Data Figure 9 | Inhibition of pannexin 1 during ultraviolet-induced apoptosis in LR73 fibroblasts promotes the formation of membrane protrusions and apoptotic bodies. **a**, ATP levels in supernatants of LR73 fibroblasts treated with 40 μM trovafloxacin with or without apoptosis induction ($n = 3$). **b**, Formation of apoptotic bodies (left) and TO-PRO-3 uptake (right) by LR73 fibroblasts treated with the indicated concentrations of trovafloxacin ($n = 3$). **c**, Generation of apoptotic bodies by LR73 fibroblasts treated with 2 mM probenecid ($n = 3$). **d**, Time-lapse images monitoring apoptotic cell morphology of LR73 fibroblasts treated with or without trovafloxacin (40 μM) or probenecid (2 mM). LR73 fibroblasts were induced to undergo apoptosis by ultraviolet treatment in all indicated experiments. Error bars in **a**–**c** represent s.e.m. Arrows, apoptopodia. Scale bars, 10 μm.



Extended Data Figure 10 | Schematic diagram depicting where pannexin 1 likely acts in limiting the fragmentation of apoptotic cells. Blocking PANX1 function (for example via trovafloxacin) leads to formation of apoptopodia, and subsequently the release of apoptotic bodies.

doi:10.1038/nature12995

The remnant of a merger between two dwarf galaxies in Andromeda II

N. C. Amorisco[1], N. W. Evans[2] & G. van de Ven[3]

Driven by gravity, massive structures like galaxies and clusters of galaxies are believed to grow continuously through hierarchical merging and accretion of smaller systems. Observational evidence of accretion events is provided by the coherent stellar streams crossing the outer haloes of massive galaxies, such as the Milky Way[1] or Andromeda[2]. At similar mass scales, around 10^{11} solar masses in stars, further evidence of merging activity is also ample[3–5]. Mergers of lower-mass galaxies are expected within the hierarchical process of galaxy formation[6], but have hitherto not been seen for galaxies with less than about 10^9 solar masses in stars[7,8]. Here we report the kinematic detection of a stellar stream in one of the satellite galaxies of Andromeda, the dwarf spheroidal Andromeda II, which has a mass of only 10^7 solar masses in stars[9]. The properties of the stream show that we are observing the remnant of a merger between two dwarf galaxies. This had a drastic influence on the dynamics of the remnant, which is now rotating around its projected major axis[10]. The stellar stream in Andromeda II illustrates the scale-free character of the formation of galaxies, down to the lowest galactic mass scales.

Andromeda II is, in size, the second-largest dwarf spheroidal galaxy known in the Local Group, with a half-light radius[11] of about 1.2 kpc (second only to Andromeda XIX[12]). With a luminosity[9] at visible wavelengths of $L_V = 9.4 \times 10^6 L_\odot$, where $L_\odot$ is the solar luminosity, it is at present 185 kpc from its host and at a heliocentric distance of about 650 kpc (ref. 13). Among the satellites of M31 (Andromeda), And II is one of the few for which a spectroscopic data set of hundreds of stars is available. These observations, made using the Deep Imaging Multi-Object Spectrograph (DEIMOS) on the Keck II telescope, revealed a strong and puzzling stellar rotation that is so far unique among the dwarf spheroidal galaxies of the Local Group[10]. The results presented here are based on a re-analysis of the latter spectroscopic observations, provided by N. Ho and M. Geha.

We assign reliable probabilities of membership to all spectroscopic targets, comprising more than 700 candidate red-giant-branch stars, by allowing for the presence of both foreground contaminants from the Milky Way and interlopers from the halo population of M31. Each of the three coexisting components (members, Milky Way halo contaminants and M31 halo contaminants) is described by a distinct spatial and kinematical distribution, the parameters of which are measured by a maximum-likelihood technique (Methods). A Bayesian approach then allows us to estimate a probability of membership for each available star: we count 632 high-probability members ($P > 0.85$; Extended Data Fig. 1).

We study the kinematical properties of And II also by a maximum-likelihood method, fully taking into account the observational uncertainties in the line-of-sight velocity of each available giant star in the spectroscopic sample. We find that, within measurement errors (median value, 6.5 km s^{-1}), there are no significant deviations from the strong rotation field of And II. However, despite such a smooth mean-velocity field and an otherwise flat velocity dispersion profile (Extended Data Fig. 2), we identify a drop and asymmetries in the velocity dispersion field, especially within the circular annulus 0.9 kpc $\lesssim R \lesssim$ 1.7 kpc.

To quantify the significance of these kinematic anomalies, we isolate a group of stars that is defined based only on its spatially connected location. Figure 1a displays the giant stars identified as high-probability members in the spectroscopic data set, superimposed on a Subaru Prime Focus Camera (Suprime-Cam) image[10] of And II. We select 134 stars over an annular, stream-like region covering an angle of 270° over the body of And II (blue points). These are compared with a control sample of 319 stars, comprising the remainder of the spectroscopic targets at comparable distances from the centre of And II (red points). Although sharing a compatible rotational field, the stream-like region is found to be kinematically colder (Fig. 1b, c). Figure 1b shows histograms of the line-of-sight velocity distribution of the available giant stars in both regions, and Fig. 1c shows the normalized generalized histograms obtained after subtracting the mean stellar rotation field. Figure 1d shows the probability distribution functions of the projected velocity dispersion, σ. The blue dashed curves in Fig. 1b–d refer to the stream-like region, and the red curves are for the control sample. The probability that the velocities of the giant stars in the two described regions have been extracted from the same parent line-of-sight velocity distribution is negligible ($P < 3 \times 10^{-6}$). This shows that, together with the stellar population of And II, an additional kinematically colder component contributes a substantial fraction of the stars in the selected annulus.

Among the stars in the latter spatially connected stream-like region, we next use a Bayesian approach to identify those that are significantly better described by the properties of the control sample with a higher velocity dispersion, $\sigma = 9.3 \pm 0.6$ km^{-1}. These are probably And II stars, and we find 14 such high-probability contaminants ($P > 0.85$; Fig. 1a, blue points). If we exclude these, the remaining 120 stars (filled blue dots) are characterized by a much lower velocity dispersion ($\sigma \lesssim 3$ km s^{-1}), the probability distribution of which is shown by the blue full curve in Fig. 1d. We identify these stars with a stream that extends coherently over a distance of 5 kpc, with an average thickness of 0.3 kpc.

The stellar stream contains at least one-tenth of the luminosity of And II, which allows us confidently to put a lower bound of $L_V \gtrsim 10^6 L_\odot$ on the total luminosity of the progenitor. Furthermore, current photometric data do not suggest that the colour spread of the red-giant-branch population of the stream is dissimilar from the considerable spread of the And II population itself (Extended Data Fig. 3). Together, these point to the progenitor of the stream being a dwarf galaxy with a total mass not too different from that of And II. A progenitor with such properties is not unexpected: recent kinematic studies of members of the Local Group have identified analogue dwarf galaxies with comparably low velocity dispersions around both the Milky Way[14] and M31[15].

The merger had a substantial influence on the dynamics and structure of the remnant. The colder stream and warmer control sample do not show statistically significant differences in the properties of their rotation, which may seem somewhat odd within the merger context. However, it is very likely that the orbital angular momentum of the merging dwarfs was much larger than the intrinsic net angular momentum of the stars in either of the dwarfs. Combined with a mass ratio expected

[1]Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen, Denmark. [2]Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK. [3]Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany.

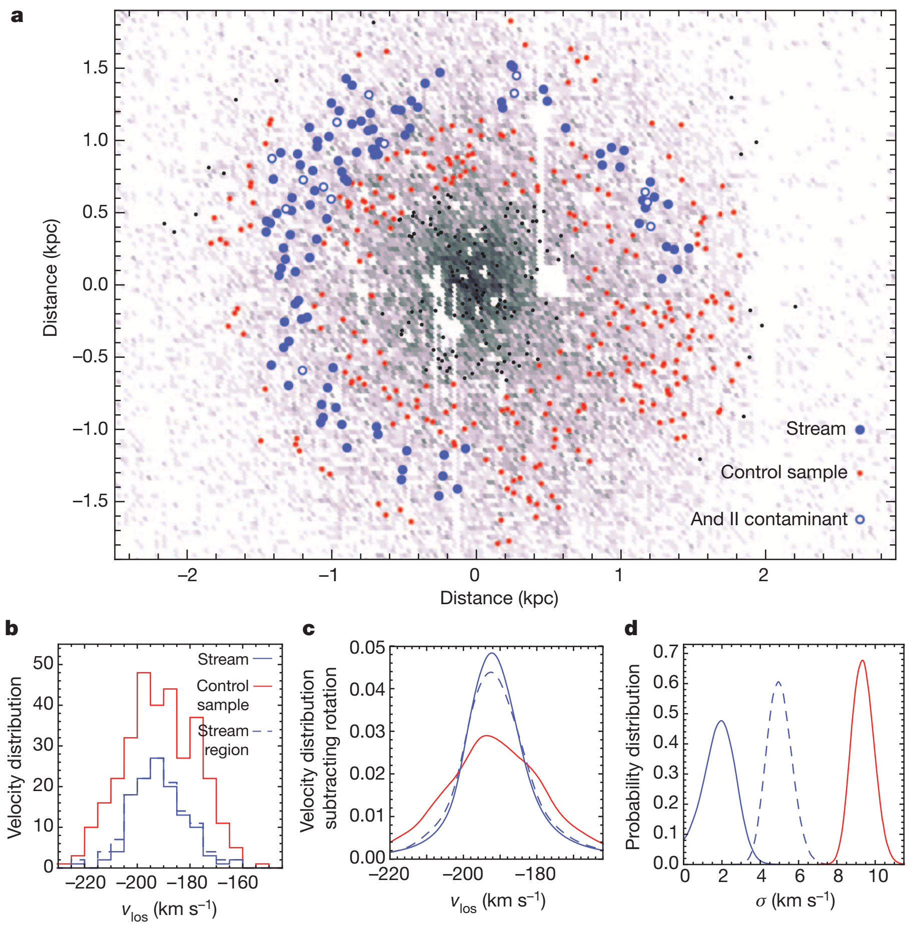


Figure 1 | Kinematic detection of a stream in And II. The targets in the circular annulus 0.7 kpc $< R <$ 1.9 kpc are divided into kinematically cold and warm components. The 134 stars in the connected stream-like region (**a**, blue points) yield the blue-dashed line-of-sight velocity, v_{los}, distributions in **b** and **c** (respectively before and after normalization, subtraction of the mean stellar rotation field and convolution with the individual measurement uncertainties), and the blue-dashed probability distribution for the projected velocity dispersion, σ, in **d**. Red distributions are associated with the kinematically warmer control sample, comprising the remaining 319 spectroscopic targets (**a**, red points). The 14 blue open points in **a** indicate stars that are more likely to belong to the main body of And II rather than to the stream itself. When they are subtracted from the sample of stream stars, the internal velocity dispersion of the stream is reduced further, leading to the blue full distributions in **b**–**d**.

to be not too far from unity, the torque exerted by the merging dwarf would have been substantial. This is most probably responsible for stars of the stream and control sample having the same rotation strength, as well as the puzzling orientation around the major axis—stellar rotation is nearly always around the minor axis, consistent with oblate axisymmetry (except for rotation around the major axis in some giant elliptical galaxies caused by triaxiality). At the same time, it is likely that more stars that once belonged to the stream's progenitor are recorded in the spectroscopic data set but cannot be clearly distinguished from the stellar population of And II because of their lower density contrast. The detection of such a stream also provides a natural explanation for the peculiar extended component of old stars[9] in And II with an effectively constant density out to a large radius of about 1.9 kpc; the merger has dynamically heated the remnant's stellar population. Although a particularly close tidal encounter[16] with M31 may also have contributed to shaping the structure of And II, this remains very uncertain given the galaxy's unknown proper motion and consequent degeneracy in the modelling[13].

We measure the line-of-sight velocity of the stream as well as its projected spatial position onto the body of And II, and use this information to constrain its approximate orbit. As Fig. 2 shows, a simple model of an orbit in a spherical potential is capable of describing the available velocity and position measurements. Although it is not possible to infer the detailed properties of the gravitational potential of And II, because the orbit is found to be almost circular, we are able to constrain the enclosed mass (stellar and dark matter) interior to the stream as follows: $M(<1.5\ \text{kpc}) = 2.5^{+3.1}_{-1.1} \times 10^8 M_\odot$. This implies a mass-to-light ratio of $45^{+60}_{-20} M_\odot/L_{\odot,\mathrm{V}}$, which is typical of dwarf spheroidal galaxies[17,18].

Even with a characteristic orbital velocity for the stream, dating the epoch of the And II merger remains challenging. The survival of cold kinematic clumps is strongly dependent on the properties of the gravitational potential in which they orbit[19,20]. Nevertheless, the And II stream seems to lie on a nearly circular orbit and never passes close to the central regions of the galaxy, which allows it to retain coherence for a very long time. At the same time, dynamical friction can drag two mutually orbiting dwarf galaxies closer, causing a merger in just a fraction of the Hubble time. We estimate that this process requires $\gtrsim$ 3 Gyr, which provides us with an approximate lower limit to associate with the merger of the And II system.

Streams of disrupted and engulfed galaxies are abundant in the halo of the Milky Way, as memorably shown in the Sloan Digital Sky Survey's 'Field of Streams'[1] and by the disrupting Sagittarius galaxy[21]. However, the frequency and role of accretion onto low-mass galaxies, and in particular of dwarf–dwarf mergers, remains unclear, given the extremely limited observational evidence. Stellar overdensities similar to those of shells have been discovered in the Fornax dwarf spheroidal galaxy[22], suggesting a late merger origin[23]. Irregular isophotes[24,25] and a kinematically cold spot[19] indicate that Ursa Minor has suffered recent disturbance, most probably the accretion of a lower-mass stellar system in the form of a stellar cluster. In this respect, And II represents a compelling case of a dwarf–dwarf merger.

Mergers between low-mass galaxies are predicted within the hierarchical framework of galaxy formation, but they are rare at present times. This is particularly true for dwarf satellite galaxies: after accretion onto their host, in this case M31, the cross-section for encounters between previously unrelated dwarf galaxies is very low[26,27], implying that subsequent merging activity is essentially limited to galaxies that were closely associated before infall. This makes the discovery of a tidal stream, originating from the engulfment of one dwarf satellite by another, particularly remarkable. As for merger events preceding infall, it is estimated[6] that one in two dark matter haloes with a virial mass of $10^{10} M_\odot$ have experienced a major merger (mass ratio $\gtrsim 1/3$) between redshift 4 and 1, but data to confirm these figures are exceedingly scarce.

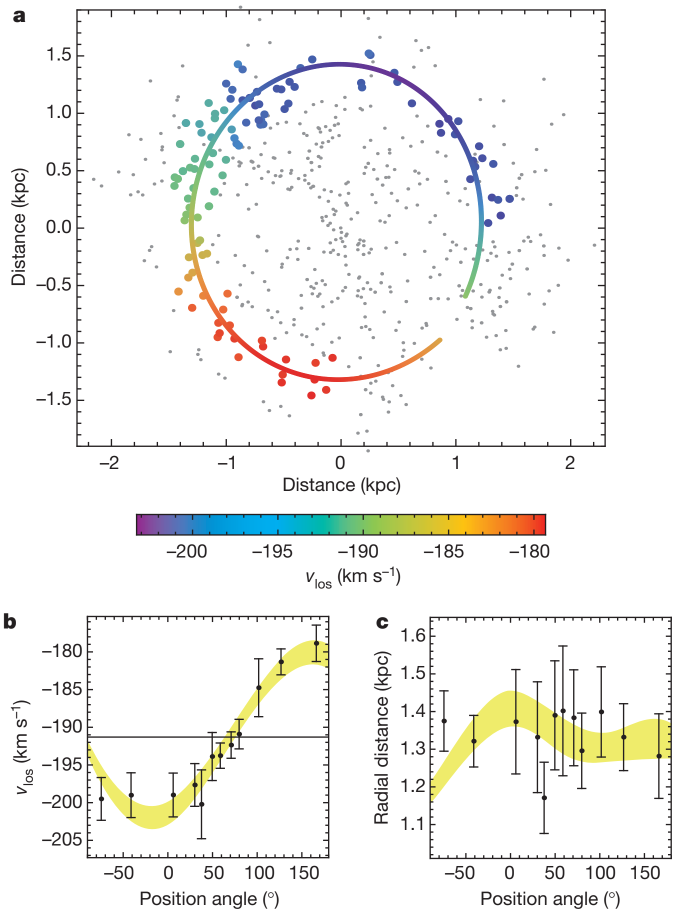


Figure 2 | The stream reproduced by an orbit in a spherical potential. **a**, Spectroscopic targets belonging to the stream, colour-coded using a smoothed velocity field for comparison with the best-fitting orbit. **b**, **c**, Line-of-sight velocity (**b**) and distance from the centre of And II (**c**) of the stars belonging to the stream, reproduced using a simple model of an orbit in a spherical potential. These data provide a direct comparison between the observables (each data point was obtained from a subset of ten stream stars; centre values, average; errors, s.d.) and the corresponding 68% confidence region obtained from the model.

Andromeda II provides direct evidence for the importance of mergers even for the smallest and least luminous of galaxies. Just as for the largest giant ellipticals, merging and accretion were dominant processes in the formation of the dwarf galaxies we see today.

METHODS SUMMARY

Spectroscopic targets that belong to the red-giant-branch population of And II are separated from Milky Way dwarf stars in the foreground and interlopers from the stellar halo of Andromeda by using maximum-likelihood and Bayesian techniques. The three distinct populations have different spatial and kinematical distributions, the properties of which are determined by the maximum-likelihood method. The extraction of the kinematics of the member stars is also obtained using a maximum-likelihood technique. This includes a full treatment of the individual observational uncertainties in the line-of-sight velocity of each spectroscopic target and allows for the subsequent kinematic detection and characterization of the stream. The luminosity estimate of the identified kinematically cold structure and inferred constraints on its progenitors are obtained by using the detailed surface density profile of And II, together with the projected region where the density contrast of the stream is highest.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 September; accepted 16 December 2013.
Published online 23 February 2014.

1. Belokurov, V. *et al.* The Field of Streams: Sagittarius and its siblings. *Astrophys. J.* **642,** L137–L140 (2006).
2. Ibata, R., Irwin, M., Lewis, G., Ferguson, A. & Tanvir, N. A giant stream of metal-rich stars in the halo of the galaxy M31. *Nature* **412,** 49–52 (2001).
3. Martínez-Delgado, D. *et al.* The ghost of a dwarf galaxy: fossils of the hierarchical formation of a nearby spiral galaxy NGC 5907. *Astrophys. J.* **689,** 184–193 (2008).
4. Chonis, Taylor S. *et al.* A petal of the sunflower: photometry of the stellar tidal stream in the halo of Messier 63 (NGC 5055). *Astron. J.* **142,** 166–181 (2011).
5. Martínez-Delgado, D. *et al.* Discovery of a giant stellar tidal stream around the disk galaxy NGC 4013. *Astrophys. J.* **692,** 955–963 (2009).
6. Fakhouri, O., Ma, C.-P. & Boylan-Kolchin, M. The merger rates and mass assembly histories of dark matter haloes in the two Millennium Simulations. *Mon. Not. R. Astron. Soc.* **406,** 2267–2278 (2010).
7. Rich, R. M, *et al.* A tidally distorted dwarf galaxy near NGC 4449. *Nature* **482,** 192–194 (2012).
8. Martínez-Delgado, D, *et al.* Dwarfs gobbling dwarfs: a stellar tidal stream around NGC 4449 and hierarchical galaxy formation on small scales. *Astrophys. J.* **748,** L24 (2012).
9. McConnachie, A. W., Arimoto, N. & Irwin, M. J. Deconstructing galaxies: a Suprime-Cam survey of Andromeda II. *Mon. Not. R. Astron. Soc.* **379,** 379–392 (2007).
10. Ho, N. *et al.* Stellar kinematics of the Andromeda II dwarf spheroidal galaxy. *Astrophys. J.* **758,** 124–136 (2012).
11. McConnachie, A. W. & Irwin, M. J. Structural properties of the M31 dwarf spheroidal galaxies. *Mon. Not. R. Astron. Soc.* **365,** 1263–1276 (2006).
12. McConnachie, A. W. *et al.* A trio of new local group galaxies with extreme properties. *Astrophys. J.* **688,** 1009–1020 (2008).
13. Watkins, L. L., Evans, N. W. & van der Ven, G. A census of orbital properties of the M31 satellites. *Mon. Not. R. Astron. Soc.* **430,** 971–985 (2013).
14. Koposov, S. E. *et al.* Accurate stellar kinematics at faint magnitudes: application to the Botes I dwarf spheroidal galaxy. *Astrophys. J.* **736,** 146 (2011).
15. Collins, M. L. M. *et al.* A kinematic study of the Andromeda dwarf spheroidal system. *Astrophys. J.* **768,** 172 (2013).
16. Shaya, E. J. & Tully, R. B. The formation of Local Group planes of galaxies. *Mon. Not. R. Astron. Soc.* **436,** 2096–2119 (2013).
17. Walker, M. in *Planets, Stars and Stellar Systems* Vol. 5 (eds Oswalt, T. D. & Gilmore, G.) 1039–1089 (Springer, 2013).
18. Amorisco, N. C. & Evans, N. W. Phase-space models of the dwarf spheroidals. *Mon. Not. R. Astron. Soc.* **411,** 2118–2136 (2011).
19. Kleyna, J. Wilkinson, M. I., Gilmore, G. & Evans, N. W. A dynamical fossil in the Ursa Minor dwarf spheroidal. *Astrophys. J.* **588,** L21–L24 (2003).
20. Sánchez-Salcedo, F. J. & Lora, V. The survival of dynamical fossils in dwarf spheroidal galaxies in conventional and modified dynamics. *Mon. Not. R. Astron. Soc.* **407,** 1135–1147 (2010).
21. Belokurov, V. *et al.* Precession of the Sagittarius stream. *Mon. Not. R. Astron. Soc.* **437,** 116–131 (2014).
22. Coleman, M. *et al.* Shell structure in the Fornax dwarf spheroidal. *Astron. J.* **127,** 832–839 (2004).
23. Amorisco, N. C. & Evans, N. W. A troublesome past: chemodynamics of the Fornax dwarf spheroidal. *Astrophys. J.* **756,** L2 (2012).
24. Irwin, M. & Hatzidimitriou, D. Structural parameters for the Galactic dwarf spheroidals. *Mon. Not. R. Astron. Soc.* **277,** 1354–1378 (1995).
25. Palma, C. *et al.* Exploring halo substructure with giant stars. IV. The extended structure of the Ursa Minor dwarf spheroidal galaxy. *Astron. J.* **125,** 1352–1372 (2003).
26. Tremaine, S. in *Structure and Evolution of Normal Galaxies* (eds Fall, S. M. & Lynden-Bell, D.) 67–84 (Cambridge Univ. Press, 1981).
27. De Rijcke, S., Dejonghe, H., Zeilinger, W. W. & Hau, G. K. T. Dwarf elliptical galaxies with kinematically decoupled cores. *Astron. Astrophys.* **426,** 53–63 (2004).

Acknowledgements We thank M. Irwin for discussions on the photometric properties of And II. The Dark Cosmology Centre is funded by the Danish National Research Foundation. This work was partly supported by Sonderforschungsbereich SFB 881 'The Milky Way System' (subproject A7) of the German Research Foundation.

Author Contributions N.C.A. performed the candidate selection using methods originally developed with N.W.E., and the subsequent kinematic extraction, together with G.v.d.V. N.C.A., N.W.E. and G.v.d.V. jointly proposed and elaborated the stream model to explain the data. The paper was written by N.C.A., with contributions from N.W.E. and G.v.d.V.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.C.A. (amorisco@dark-cosmology.dk).

METHODS

Membership selection. To address the kinematical properties of And II, we need to separate the spectroscopic targets that belong to the red-giant-branch population of the dwarf galaxy from any contaminants. Contamination has two different origins: dwarf stars in the foreground belonging to the Milky Way and interlopers from the stellar halo of Andromeda. To disentangle these three distinct components reliably, we model the spectroscopic data set as a superposition of multiple independent stellar populations, within the framework of a maximum-likelihood technique[23,28]:

$$L(\Theta)=\prod_j\sum_i f_i P^{\rm sp}\left(X_j;\Theta_i^{\rm sp}\right)P^{\rm kin}\left(v_j;\Theta_i^{\rm kin}\right)$$

The index j runs over the spectroscopic targets, and i indicates the three populations of the model, each containing a fraction of stars f_i, with the constraint $\sum_i f_i=1$. The probability function $P_i^{\rm sp}$, parameterized by the set of parameters $\Theta_i^{\rm sp}$, describes the spatial distribution of the members of the component i on the plane of the sky. Given the limited angular size of And II, we can adopt a constant surface density distribution for both populations of contaminants, and use a Plummer density profile with elliptical isophotes to describe the population of And II itself. Analogously, the three components have different probability distributions for their kinematics, which are respectively described by the functions $P_i^{\rm kin}$ and the parameters $\Theta_i^{\rm kin}$. Given the significant separation of the contaminants from And II in terms of systematic velocity[10], we can describe both the Milky Way and the Andromeda halo population with a simple Gaussian line-of-sight velocity distribution. The observational uncertainty in the measurement of the line-of-sight velocity of each single spectroscopic target, δ_j, is fully included in the analysis by formal convolution; hence, for the contaminants, we have

$$P_{\rm cont}^{\rm kin}\left(v_j;v_i^{\rm sys},\sigma_i\right)=\frac{1}{\sqrt{2\pi\left(\sigma_i^2+\delta_j^2\right)}}\exp\left[-\frac{1}{2}\frac{\left(v_j-v_i^{\rm sys}\right)^2}{\sigma_i^2+\delta_j^2}\right]$$

where Milky Way and Andromeda contaminants have their own systematic velocity, $v_i^{\rm sys}$, and intrinsic velocity dispersion, σ_i. The velocity distribution of And II is also normally distributed, but has a rotating mean velocity field, with frequency Ω:

$$P_{\rm And\,II}^{\rm kin}\left(v_j;v^{\rm sys},\Omega,\sigma\right)=\frac{1}{\sqrt{2\pi\left(\sigma^2+\delta_j^2\right)}}\exp\left[-\frac{1}{2}\frac{\left(v_j-v^{\rm sys}-\Omega\bullet X_j\right)^2}{\sigma^2+\delta_j^2}\right]\quad(1)$$

The parameter space is explored by means of a suite of Monte Carlo chains, constructed using the Metropolis–Hastings algorithm[29]. Once the best-fitting model is identified, for example as described by the set of parameters $\Theta^{\rm bf}$, a Bayesian approach allows us to estimate, for each spectroscopic target j, the probability of that individual star belonging to the population i:

$$p_j^i=\frac{f_iP^{\rm sp}\left(X_j;\Theta_i^{\rm bf,sp}\right)P^{\rm kin}\left(v_j;\Theta_i^{\rm bf,kin}\right)}{\sum_k f_kP^{\rm sp}\left(X_j;\Theta_k^{\rm bf,sp}\right)P^{\rm kin}\left(v_j;\Theta_k^{\rm bf,kin}\right)}$$

Extended Data Fig. 1 shows the distribution of all spectroscopic targets in the plane ($v_{\rm los}$, R), where R is the projected radial distance from the centre of And II. Colour-coding of each target is set by its probability of membership of And II: the combination of spatial position and kinematical information is capable of efficiently selecting the members of And II. As a further check, we also exclude a few targets that appear as fortuitous 'kinematic members', but whose membership to And II is questioned[10] either by their colour ($V-I>2.5$) or by the strength of the Na I line ($EW_{\rm Na\,I}>4$), which indicate that these may in fact be dwarf foreground stars from the Milky Way halo. As a result, we identify 632 giant stars belonging to And II with high probability ($P>0.85$).

Kinematic extraction. Despite significant observational effort and use of the DEIMOS spectrograph on Keck, the heliocentric distance of And II, $D\approx650$ kpc, is so great that uncertainties in the measurement of the line-of-sight velocity of each target are significantly higher than figures achieved for the closer Milky Way dwarf satellites ($\delta_v\approx2\,{\rm km\,s^{-1}}$). In particular, with a median of about $\delta_v\approx6.5\,{\rm km\,s^{-1}}$, measurement errors are comparable in magnitude to the kinematical spread due to the intrinsic velocity dispersion of the dwarf, which implies that appropriate treatment of these uncertainties is crucial.

For this reason, we extract the kinematic properties of And II by fully taking into account all measurement errors, individually for each spectroscopic target. First, we adopt the line-of-sight velocity distribution of equation (1) to infer the intrinsic velocity dispersion, σ, in circular annuli as displayed in Extended Data Fig. 2. The resulting σ profile is flat, as is typically observed in dwarf spheroidal galaxies, except for a significant dip around $R\approx1.3$ kpc, independently of the precise choice of the circular annuli.

Next we applied the same technique to stars in the circular annulus with 0.9 kpc $\lesssim R\lesssim1.7$ kpc, but now allowing for two spatially connected components with a different intrinsic velocity dispersion. This results in the identification of the stream-like region indicated by blue circles in Fig. 1a. The convolution of the intrinsic line-of-sight velocity distribution with the error function of each target, which is assumed to be normally distributed, is illustrated in Fig. 1c, where we show normalized generalized histograms for the velocity distributions of both the stream and the control sample. Only by explicitly including each measurement error individually is it then possible to infer the probability distributions of the intrinsic dispersion for the stream and the control sample (Fig. 1d). Well over 10^5 draws in a suite of Monte Carlo chains have been used to sample the tails of these σ distributions accurately. This technique also ensures that these σ distributions are properly marginalized against uncertainties in all other parameters of the model.

Stream luminosity and progenitor. Our kinematical analysis identifies $N_{\rm str}=120$ red giant stars as high-probability members of the kinematically cold stream, in a pool of $N_{\rm mem}=632$ members of And II. Although indicative of the significant luminosity of this structure, it is not entirely correct to use simply the ratio $N_{\rm str}/N_{\rm mem}$ to derive an estimate of the luminosity of the stream itself. This is because the stream covers a limited area over the body of And II and the spectroscopic coverage is neither uniform over the body of the dwarf nor proportional to its surface brightness profile.

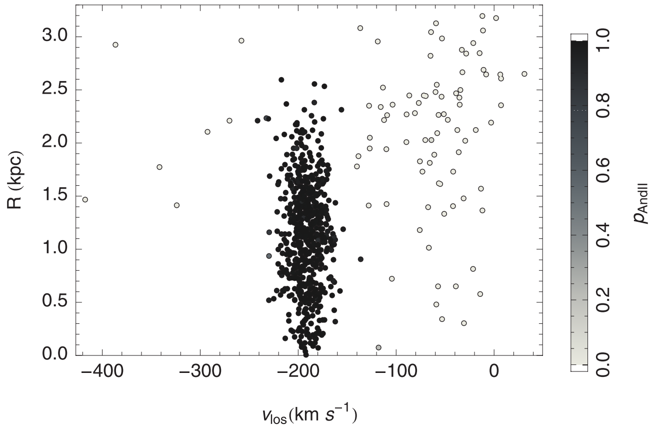
We can instead obtain more accurate insight into the luminosity of the stream by restricting ourselves to the projected region it covers. This has been identified in Fig. 1a as an approximately annular region, centred around $R\approx1.35$ kpc and covering an angle of $\sim270^\circ$. By using the detailed surface density profile[9] of And II, obtained by using deep Suprime-Cam data, we calculate that the projected area contains ~15% of the total luminosity of the dwarf. In turn, we find only 14 high-probability contaminants in this region, that is, stars that are significantly more likely to belong to the And II population than to the kinematically cold stream. This implies a luminosity estimate for the stream of

$$L_{\rm str,V}\approx0.15\times\frac{120}{134}L_{\rm And\,II,\,V}\approx0.13L_{\rm And\,II,\,V}\approx1.3\times10^6L_\odot\quad(2)$$

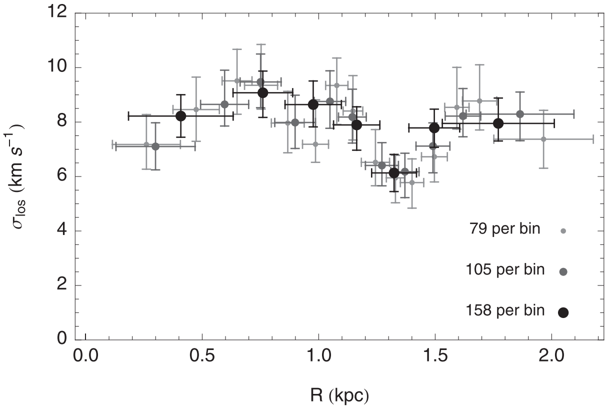
It is worth mentioning that, on the basis of kinematical data alone, we are inevitably capable of identifying only those parts of the stream that have a sufficiently high density contrast relative to the And II population, and whose kinematics in the stream's progenitor were cold enough to stand out against the average present properties of And II. In particular, as a result of the merger, it is plausible that any stellar components that were originally more diffuse in the stream's progenitor are now dispersed in And II, eluding purely kinematical detection. This implies that the above 13% (equation (2)) should be regarded as a lower limit for the luminosity of the progenitor itself.

Whereas the resulting luminosity already makes a stellar cluster origin unlikely, this can be excluded further by considering the properties of the distribution of the stream's members in the colour–magnitude diagram. Extended Data Fig. 3 follows the same stream versus control sample labelling as in Fig. 1a, to allow for the comparison of the properties of the stream's member stars with those of the And II population. We find that this data suggest little difference in the distribution of such two populations in the colour–magnitude diagram, highlighting the fact that the stream also has a significant colour spread.

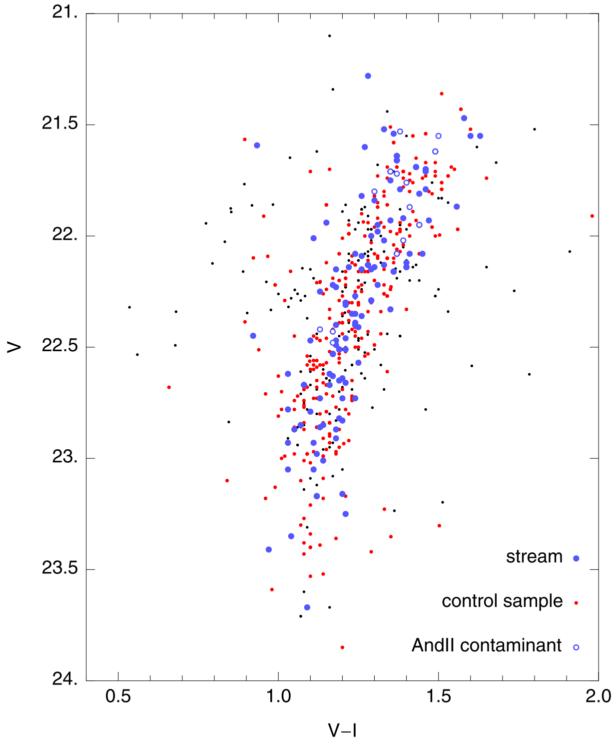
28. Walker, M. G. & Peñarrubia, J. A method for measuring (slopes of) the mass profiles of dwarf spheroidal galaxies. *Astrophys. J.* **742,** 20 (2011).
29. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21,** 1087–1092 (1953).



Extended Data Figure 1 | Membership selection. The spectroscopic data set in the plane (v_{los}, R), shaded according to the probability of each target belonging to the stellar population of And II. Non-member targets with velocities higher than the systematic velocity of And II, $v_{sys} = -191.4 \pm 0.4$, are foreground contaminants from the Milky Way, whereas non-member targets at lower negative velocities are interlopers from the Andromeda halo.



Extended Data Figure 2 | Velocity dispersion profile. Andromeda II has an approximately flat velocity dispersion profile, except for a significant dip near the average projected radius of the stellar stream. Points of different sizes and shading depths refer to different circular annuli sizes (as in the key), and error bars display 68% confidence levels around the most likely central value.



Extended Data Figure 3 | Colour–magnitude diagram. The distribution of the stars belonging to the stream (blue points) and stars in the control sample (red points) in V-band magnitude versus $V - I$ colour.

doi:10.1038/nature13046

Rotationally driven 'zebra stripes' in Earth's inner radiation belt

A. Y. Ukhorskiy[1], M. I. Sitnov[1], D. G. Mitchell[1], K. Takahashi[1], L. J. Lanzerotti[2] & B. H. Mauk[1]

Structured features on top of nominally smooth distributions of radiation-belt particles at Earth have been previously associated with particle acceleration and transport mechanisms powered exclusively by enhanced solar-wind activity[1–4]. Although planetary rotation is considered to be important for particle acceleration at Jupiter and Saturn[5–9], the electric field produced in the inner magnetosphere by Earth's rotation can change the velocity of trapped particles by only about 1–2 kilometres per second, so rotation has been thought inconsequential for radiation-belt electrons with velocities of about 100,000 kilometres per second. Here we report that the distributions of energetic electrons across the entire spatial extent of Earth's inner radiation belt are organized in regular, highly structured and unexpected 'zebra stripes', even when the solar-wind activity is low. Modelling reveals that the patterns are produced by Earth's rotation. Radiation-belt electrons are trapped in Earth's dipole-like magnetic field, where they undergo slow longitudinal drift motion around the planet because of the gradient and curvature of the magnetic field. Earth's rotation induces global diurnal variations of magnetic and electric fields that resonantly interact with electrons whose drift period is close to 24 hours, modifying electron fluxes over a broad energy range into regular patterns composed of multiple stripes extending over the entire span of the inner radiation belt.

Radiation belts comprise electrons and ions trapped in the near-Earth region (distance from the Earth's centre $\lesssim 6$ Earth radii, R_{E}) where the planetary magnetic field is approximately dipolar. The energy range of trapped particles spans multiple orders of magnitude, from about 10 kilo-electron volts (keV) to multiple tens of mega-electron volts (MeV), comprising respectively energetic particles, and relativistic electrons and penetrating ions[1]. The electron population is divided into the inner and the outer belts by a so-called slot region where particle loss mechanisms dominate and the electron fluxes are depleted[10,11].

Measurements from the RBSPICE instruments[12] on board the twin Van Allen Probes spacecraft mission[13] show that the lower-energy electrons (<1 MeV) across the entire inner radiation belt ($\lesssim 3\,R_{\mathrm{E}}$) are organized in regular patterns, which we refer to as 'zebra stripes' (Fig. 1, Extended Data Fig. 1). Similar patterns were previously inferred from low-altitude electron measurements and interpreted as a manifestation of storm-time geomagnetic activity[14]. It was suggested that the patterns are produced by a resonant interaction of energetic electrons with monochromatic ultralow-frequency (ULF; ~1 mHz) perturbations of Earth's magnetic field related to enhanced solar-wind driving, and that in order to create multiple zebra stripes the perturbations must be composed of multiple wave modes with discrete spatial structure.

From the analysis of a continuous 6-month interval of the RBSPICE measurements, we conclude that these zebra stripes are ubiquitous features in the inner radiation belt. The patterns were observed during all spacecraft orbits and appeared to be more regular during quiet geomagnetic conditions (quiet times) than during geomagnetic storms (Fig. 1). Considering the ULF hypothesis, we note that during average geomagnetic conditions global monochromatic ~1 mHz perturbations in the magnetic field are not typically observed at inner-belt altitudes. The perturbations in this frequency range are directly driven by variations in the solar-wind flow[15,16], and these variations have broad-band turbulent spectra[17]. The disturbances propagate into the inner belt as magnetosonic waves that typically do not exhibit discrete spatial structure. A new explanation is needed for the regular organization of the inner-belt electrons that must be viable for both quiet and active geomagnetic conditions in the absence of monochromatic ULF wave activity.

Trapped electrons undergo slow longitudinal drift motion around Earth as a consequence of the gradient and curvature of its nearly dipole magnetic field[18]. The drift frequency, ω_{D}, depends on the electron energy, K, and the location measured by L (the radial distance in R_{E} in the equatorial plane) as $\omega_{\mathrm{D}} \propto KL$. The energy along the individual zebra stripes in the electron energy spectra depends on radial distance as $K \propto 1/L$, which means that the electron drift frequency along each stripe is constant. This is further supported by the distribution of particle pitch angles (the angle between a particle's velocity and the direction of the magnetic field) at constant energy values along the stripes. The distributions are curved towards larger L values for more field-aligned particles, following the dependence of the drift frequency on the equatorial pitch angle[1] (see Methods and Extended Data Fig. 1). The organization of stripes along constant values of the longitudinal drift frequency suggests that the zebra stripes are produced in a global process that modulates the electron longitudinal drift motion across the entire inner belt.

The only global mechanism in the inner magnetosphere that is independent of solar-wind activity is the rotation of Earth. The rotation induces an electric field that controls cold magnetospheric plasma within a $\lesssim 4\,R_{\mathrm{E}}$ region around Earth, referred to as the co-rotation region (Fig. 2). Outside the co-rotation region, the plasma is controlled by the electric field produced by the interaction of the solar-wind flow with the planetary magnetic field[19,20]. Owing to an $\sim 11^{\circ}$ offset between the magnetic dipole axis and the planetary rotation axis, the entire co-rotation region oscillates relative to the outer magnetosphere with a period of 24 h. This relative motion causes global diurnal perturbations of the magnetic field over the entire co-rotation region. Because the inner-belt electrons are embedded inside the co-rotation region, their drift motion around Earth is influenced by the inductive electric field produced by these diurnal perturbations.

To understand how the diurnal perturbations can affect the distribution of energetic electrons in the inner belt, we used a test-particle model[21] (see Methods for details). Because the longitudinal drift motion of the inner-belt electrons is well decoupled from their gyro-motion and bounce motion[1], test-particle trajectories were computed in the guiding centre approximation (see Methods) and consideration was restricted to particles bouncing at the magnetic equator. Electrons with initially smooth energy and radial distributions drifting in a dipole magnetic field were subjected to global diurnal oscillations of the electric field. By slowly varying the energy and radial position of individual particles, the electric field gradually modified the electron distribution into a pattern of zebra stripes (Fig. 1d and Supplementary Video 1) that is remarkably similar to observations (Fig. 1a–c).

The model shows that the single monochromatic diurnal perturbation of the electric field produces a pattern of multiple seemingly regular

[1]Applied Physics Laboratory, Johns Hopkins University, 11100 Johns Hopkins Rd, Laurel, Maryland 20723, USA. [2]New Jersey Institute of Technology, 141 Summit St, Newark, New Jersey 07103, USA.

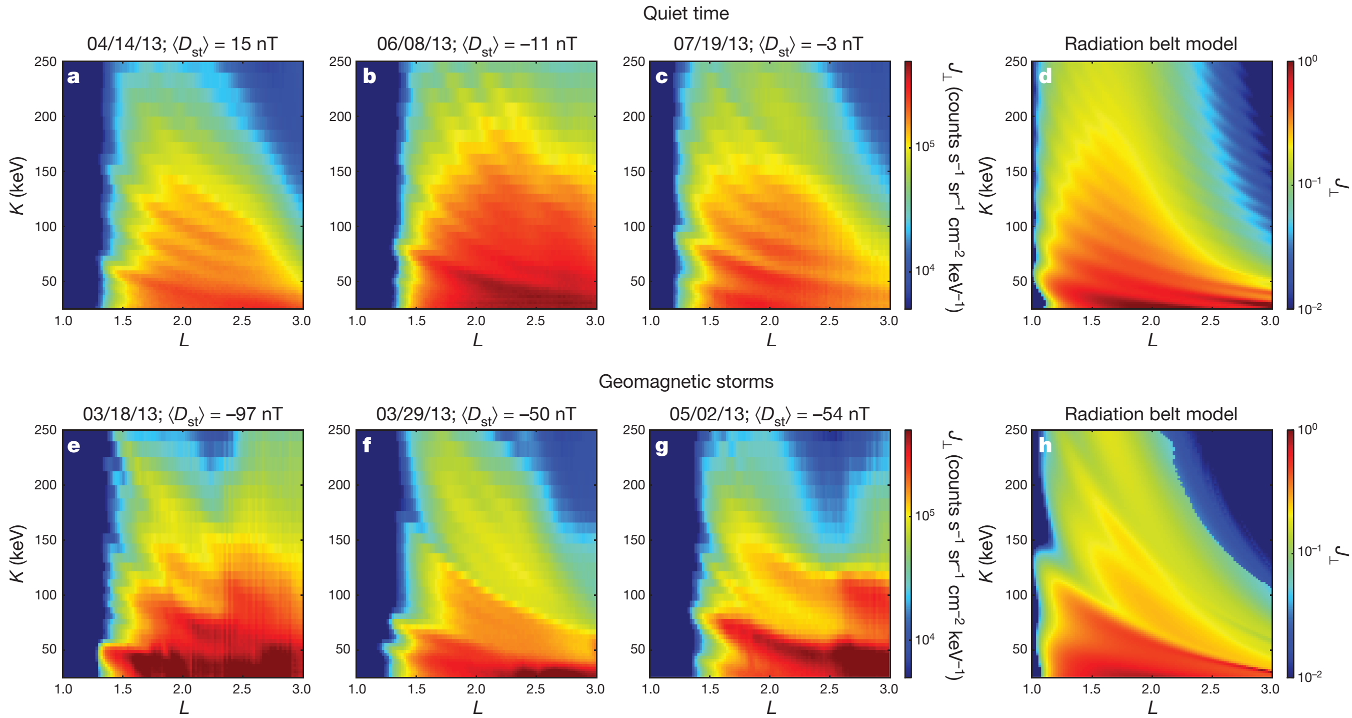


Figure 1 | Zebra stripe patterns in energetic electron distributions from the inner radiation belt. a–g, Examples of energetic electron spectra measured by RBSPICE A, averaged over 45 s, in the inner radiation belt during quiet times (**a–c**) and during geomagnetic storms (**e–g**), parameterized by the 9-h averaged geomagnetic activity ⟨D_{st}⟩ index, compared with simulated spectra from a test-particle radiation-belt model (**d** and **h**).

zebra stripes that extends over a broad energy range (~20–250 keV). This result is surprising, because it was previously believed that multiple stripes could be formed only if the perturbation spectrum consisted of multiple resonant harmonics[14]. The physical mechanism of zebra stripe production is illustrated in Fig. 3. The electron drift motion conserves the first adiabatic invariant (the magnetic moment, $\mu \propto KL^3$) associated with the fast gyration of particles perpendicular to the magnetic field[18]. The change in electron energy is therefore uniquely related to the change in its radial position. The phase space of the electron drift motion at a given value of the first invariant is specified by its energy K and its longitude, measured by the azimuthal angle φ. In the absence of perturbations in the field, electron energy is conserved and electrons drift on circular orbits at constant frequencies defined by their energy. A monochromatic diurnal perturbation in the electric field modifies the phase space motion. Particles with a drift period of ~24 h are trapped at resonance with the perturbation in the field. Their drift phases and energies oscillate about the resonance values, forming a phase space island bounded by a separatrix. The phase space trajectories outside the separatrix correspond to non-resonant particles. Over the course of its drift period a non-resonant particle has to flow around the resonant island, which involves a variation in its energy and a change in the drift frequency. The magnitude of the change in energy and the instantaneous drift frequency depend on the particle's proximity to the separatrix. Consequently, non-resonant electrons with initially identical values of energy and radial position but uniformly distributed in the drift phase undergo phase mixing, which has a similar effect to the 'stretching and folding' found in chaotic maps[22] (Fig. 3b, Supplementary Video 2, Extended Data Fig. 2).

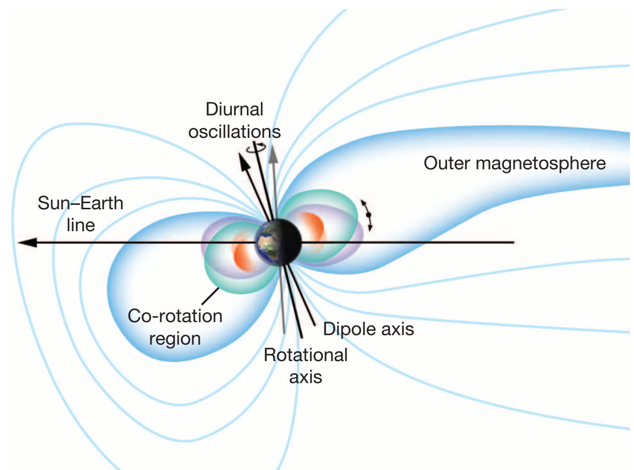


Figure 2 | Global diurnal oscillations of the electric and magnetic fields produced by Earth's rotation. Owing to an offset between Earth's rotation axis and the magnetic dipole axis, the co-rotation region (green at one time, purple 12 h later) oscillates relative to the outer magnetosphere (blue) with a period of 24 h. This relative motion of two domains of the magnetized plasma (indicated by a black double-headed arrow) results in diurnal perturbations of the magnetic and the inductive electric fields over the entire co-rotation region that affect the inner-belt electrons (orange) embedded in the co-rotation region.

The folds in the electron distribution function produced by the phase mixing process are manifested as zebra stripes in the energy spectra. The spacing between the phase space folds changes with time, which can cause variations in the zebra patterns observed during quiet-time conditions (Fig. 1a–c, Extended Data Fig. 3). Stretching and folding leads to a continuous increase in the density of stripes in the zebra patterns. If the process could continue uninterrupted, a zebra pattern monitored with a detector of finite energy resolution would eventually merge into a seemingly continuous spectrum. The ubiquitous presence of the zebra patterns at all analysed spacecraft orbits suggests that the stretching and folding process is occasionally disrupted by additional perturbations, which reset energetic electron distributions to new initial conditions.

During geomagnetic storms, the co-rotation region and the inner belt are affected by additional perturbations driven by enhanced solar-wind activity. To understand how that may affect the zebra patterns, we considered one possible source of such perturbations, namely, magnetospheric substorms. Substorms involve global reconfigurations of Earth's magnetotail leading to disturbances that can propagate deep into the

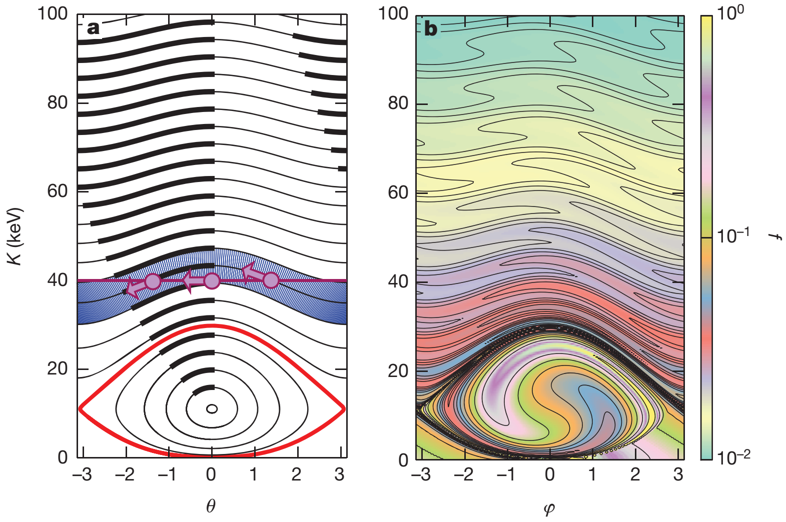


Figure 3 | Formation of zebra stripes. **a**, The phase space portrait of electron motion at a constant value of the first adiabatic invariant ($K = 30$ keV at $L = 2$) in the frame moving at the perturbation phase velocity, θ ($= \varphi - \omega_0 t$). The separatrix (red) divides resonant and non-resonant trajectories. The bold segments show the intervals traversed over a fixed time interval at different initial conditions. Particles (filled pink circles with arrows) initially distributed over a line segment (purple) eventually spread over the phase space region (blue). **b**, A snapshot of the zebra pattern produced from an initially smooth distribution. The phase space density value (f) is shown with colour.

inner magnetosphere. We modified our test-particle simulations by adding a perturbation electric field based on a statistical distribution of substorm recurrence frequency[23]. The model results (Fig. 1h and Supplementary Videos 3 and 4) are qualitatively similar to energetic electron distributions observed during geomagnetic storms (Fig. 1e–g). We therefore conclude that although patterns of zebra stripes in Earth's trapped electrons are observed at any level of solar-wind activity, enhanced activity can substantially modify and even disrupt their structure. Additional work is in progress to understand the potential role of ULF waves during enhanced geomagnetic activity in the modification of zebra stripes.

METHODS SUMMARY

To understand how zebra patterns can be produced by global diurnal oscillations of the inductive electric field associated with Earth's rotation, we used a test-particle model that computes the guiding centre motion of equatorially mirroring electrons in Earth's dipole magnetic field. Evolution of an initially smooth distribution of the electron phase space density, $f(L, K) = A(1 - e^{-\alpha L})K^{-\beta(L)}$, where A is a normalization constant and the parameters $\alpha = 5$ and $\beta(L) = -1.75 - 0.80(L - 1)^2$ were chosen to represent average observed electron distributions that we estimated for the region between $L = 1$ and $L = 3$ and $K = 25$ keV and $K = 250$ keV, was computed with the use of 1.3×10^4 test particles. The dominant effect of diurnal oscillations in the field was simulated with an azimuthal component of the inductive electric field $E_\varphi = E_0 \sin(\varphi - \omega_0 t + \psi_0)$, where the amplitude $E_0 = 0.2\,\mathrm{mV\,m^{-1}}$, φ is the azimuthal drift angle (longitude), the frequency $\omega_0 = 2\pi/T_0$, the period $T_0 = 24$ h, and ψ_0 is a random phase shift. The effect of storm-time magnetospheric substorms was simulated by adding a perturbation electric field $E_{\varphi 1} = E_1 \sum_i \sin(\varphi - \omega_i t + \psi_i)$, $i = 1,...,5$, where $E_1 = 0.32\,\mathrm{mV\,m^{-1}}$, $\omega_i = 2\pi/T_i$ and $T_i = 2$–4 h corresponding to the most probable recurrence of magnetospheric substorms[23].

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 October 2013; accepted 16 January 2014.

1. Schulz, M. & Lanzerotti, L. J. *Physics and Chemistry in Space* Vol. 7, *Particle Diffusion in the Radiation Belts* (ed. Roederer, J. G.) (Springer, 1974).
2. Horne, R. B. Plasma astrophysics: acceleration of killer electrons. *Nature Phys.* **3,** 590–591 (2007).
3. Hudson, M. K., Kress, B. T., Mueller, H.-R., Zastrow, J. A. & Blake, J. B. Relationship of the Van Allen radiation belts to solar wind drivers. *J. Atmos. Sol. Terr. Phys.* **70,** 708–729 (2008).
4. Thorne, R. M. Radiation belt dynamics: the importance of wave-particle interactions. *Geophys. Res. Lett.* **37,** L22107 (2010).
5. Brice, N. M. & Ioannidis, G. A. The magnetospheres of Jupiter and Earth. *Icarus* **13,** 173–183 (1970).
6. Hill, T. W. Inertial limit on corotation. *J. Geophys. Res.* **84,** 6554–6558 (1979).
7. Vasyliūnas, V. M. in *Physics of the Jovian Magnetosphere* (ed. Dessler, A. J.) 395–453 (Cambridge Univ. Press, 1983).
8. Cowley, S. W. H., Bunce, E. J., Stallard, T. S. & Miller, S. Jupiter's polar ionospheric flows: theoretical interpretation. *Geophys. Res. Lett.* **30,** 1220–1224 (2003).
9. Mauk, B. H. *et al.* in *Saturn from Cassini-Huygens* (eds Dougherty, M. K., Esposito, L. W. & Krimigis, S. M.) 281–331 (Springer, 2009).
10. Lyons, L. R., Thorne, R. M. & Kennel, C. F. Pitch-angle diffusion of radiation belt electrons within the plasmasphere. *J. Geophys. Res.* **77,** 3455–3474 (1972).
11. Meredith, N. P. *et al.* Relativistic electron loss timescales in the slot region. *J. Geophys. Res.* **114,** A03222 (2009).
12. Mitchell, D. G. *et al.* Radiation Belt Storm Probes Ion Composition Experiment (RBSPICE). *Space Sci. Rev.* **179,** 263–308 (2013).
13. Mauk, B. H. *et al.* Science objectives and rationale for the Radiation Belt Storm Probes mission. *Space Sci. Rev.* **179,** 3–27 (2013).
14. Sauvaud, J.-A. *et al.* Inner radiation belt particle acceleration and energy structuring by drift resonance with ULF waves during geomagnetic storms. *J. Geophys. Res.* **118,** 1–14 (2013).
15. Matsuoka, H., Takahashi, K., Yumoto, K., Anderson, B. J. & Sibeck, D. G. Observation and modeling of compressional Pi 3 magnetic pulsations. *J. Geophys. Res.* **100,** 12103–12115 (1995).
16. Kessel, R. L. Solar wind excitation of Pc5 fluctuations in the magnetosphere and on the ground. *J. Geophys. Res.* **113,** A04202 (2008).
17. Petrosyan, A. *et al.* Turbulence in the solar atmosphere and solar wind. *Space Sci. Rev.* **156,** 135–238 (2010).
18. Ukhorskiy, A. Y. & Sitnov, M. I. Dynamics of radiation belt particles. *Space Sci. Rev.* **179,** 545–578 (2013).
19. Nishida, A. Formation of plasmapause, or magnetospheric plasma knee, by the combined action of magnetospheric convection and plasma escape from the tail. *J. Geophys. Res.* **71,** 5669–5679 (1966).
20. Wolf, R. A., Spiro, R. W., Sazykin, S. & Toffoletto, F. R. How the Earth's inner magnetosphere works: an evolving picture. *J. Atmos. Sol. Terr. Phys.* **69,** 288–302 (2007).
21. Ukhorskiy, A. Y. & Sitnov, M. I. Radial transport in the outer radiation belt due to global magnetospheric compressions. *J. Atmos. Sol. Terr. Phys.* **70,** 1714–1726 (2008).
22. Smale, S. Differentiable dynamical systems. *Bull. Am. Math. Soc.* **73,** 747–817 (1967).
23. Borovsky, J. E., Nemzek, R. J. & Belian, R. D. The occurrence rate of magnetospheric-substorm onsets: random and periodic substorms. *J. Geophys. Res.* **98,** 3807–3814 (1993).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by NSF grant AGS1059736, NASA grant NNX11AO74G and NASA contract NAS5-01072 through a subcontract from NJIT 999640-I. We thank A. Marcotte (JHU/APL) for help with Fig. 2.

Author Contributions A.Y.U. was lead author, developed the hypothesis of the rotationally driven nature of the zebra pattern and was responsible for all theoretical analysis. M.I.S. contributed to theoretical analysis and simulations. D.G.M. was responsible for the RBSPICE data analysis. K.T. provided consultations on the ULF wave–particle interaction. L.J.L. and B.H.M. participated in data analysis and in the interpretation of theoretical work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.Y.U. (ukhorskiy@jhuapl.edu).

METHODS

Energetic electrons trapped in the approximately dipolar magnetic field in the inner belt exhibit quasiperiodic motion, which consists of three components: rapid gyromotion around the guiding centre, the bounce motion of particle guiding centre along the field lines between conjugate reflection points, and slow gradient-curvature longitudinal drift of the guiding centre around Earth. Each component of the electron motion is associated with an adiabatic invariant[24].

Electron energy along the individual stripes of zebra patterns observed by RBSPICE varies as $1/L$, which means that the stripes trace constant values of the gradient-curvature drift frequency. This, along with the large spatial extent of the stripes, suggests that their formation involves a global mechanism that interacts with the electron gradient-curvature drift motion associated with the third adiabatic invariant. An example of zebra stripes is shown in Extended Data Fig. 1. The lower panel shows the energy spectrum of energetic electrons with approximately 90° pitch angles along a perigee pass of the Van Allen Probes A. The pitch-angle distribution of $\sim$230 keV electrons along the pattern is shown in the upper panel of the figure. Electrons have a highly anisotropic pitch-angle distribution centred at 90°; a fit with the canonical $\sin^n\alpha$ distribution (where α is the pitch angle) yields $n = 4$. Zebra stripes extend across the entire pitch-angle distribution and are curved towards larger L values for more field-aligned particles. The curved shape of the stripes reflects the pitch-angle and L dependence of the bounce-averaged gradient-curvature drift frequency[1] which further points to the fact that the zebra patterns are formed in a global process that modifies the electron drift motion.

To understand how zebra patterns can be produced by global diurnal oscillations of the inductive electric field associated with Earth's rotation, we used a test-particle model. Because the spatial and temporal scales of the gradient-curvature drift motion of energetic electrons in the inner belt are separated from the spatial and temporal scales of their gyro- and bounce motions by orders of magnitude[1], the formation of zebra stripes does not violate either the first or the second adiabatic invariants of the electron motion. From the conservation of the first invariant it follows that the electron motion can be described in the guiding centre approximation[25,26]. Because the stripes are observed at all values of the narrow pitch-angle distribution of trapped particles, the mechanism can be illustrated for a fixed value of the second invariant. We therefore restrict our consideration to the equatorially mirroring particles for which the pitch-angle distribution peaks.

In a dipole magnetic field, the guiding centre drift due to the magnetic field gradient and curvature is in the azimuthal direction. A small-amplitude (that is, $E \ll B$) electric field causes an additional drift, $\boldsymbol{u}_E \sim \boldsymbol{E}\times\boldsymbol{B}$, which drives particle motion in the radial direction but does not considerably affect the azimuthal motion dominated by the gradient drift. In the presence of monochromatic oscillations in the electric field produced by Earth's rotation, the guiding centre motion of equatorially mirroring particles is defined by:

$$\begin{cases} \dot{L} = \dfrac{cE_0}{R_E B_0} L^3 \sin(\varphi - \omega_0 t + \psi) \\ \dot{\varphi} = \omega_D(L) = \dfrac{\mu c}{\gamma e} \dfrac{1}{(R_E L)^2} \end{cases} \qquad (1)$$

where E_0 is the electric field amplitude, $\omega_0 = 2\pi/T_0$ is the electric field frequency, $T_0 = 24$ h, φ is the azimuthal angle, ψ is a phase shift, ω_D is the frequency of the gradient and curvature drift motion, μ is the first adiabatic invariant, γ is the relativistic factor, c is the speed of light, $B_0 = 3.11 \times 10^4$ nT is magnetic field intensity on Earth's surface at the Equator, and $R_E = 6.4 \times 10^3$ km is Earth's radius.

Equations (1) were solved numerically with $E_0 = 0.2\ \mathrm{mV\,m^{-1}}$ for an ensemble of 1.3×10^4 test particles initially evenly distributed between $L = 1$ and $L = 3$ with energies between $K = 25$ keV and $K = 250$ keV. Global evolution of the energetic electron distributions across the inner belt was computed using Liouville's theorem: the values of electron phase density were traced along the calculated test-particle trajectories backwards in time to the initially smooth distribution:

$$f(L,K) = A(1 - e^{-\alpha L})K^{-\beta(L)} \qquad (2)$$

where the parameters $\alpha = 5$ and $\beta(L) = -1.75 - 0.80(L-1)^2$ were chosen to fit the averaged distribution of energetic electrons measured by RBSPICE A. Numerical simulations (Fig. 1d and Supplementary Video 1) showed the emergence of zebra patterns that are remarkably similar to the patterns observed by RBSPICE A during quiet-time conditions (Fig. 1a–c).

The mechanism of zebra stripe formation is illustrated in Fig. 3, Supplementary Video 2 and Extended Data Fig. 2. Figure 3a shows the phase portrait of equations (1) along the surface of a constant value of the first adiabatic invariant: $L \propto \mu K^{-1/3}$ ($K = 30$ keV at $L = 2$). The system has two different types of solution. Resonant particles with a drift period of $\sim$24 h are trapped in the potential well formed by diurnal oscillations in the electric field. Their azimuthal angle and energy oscillate about the resonance values forming a phase space island bounded by a separatrix. The frequency of these nonlinear oscillations depends on particle proximity to the resonance which causes the phase mixing: particles with similar initial conditions gradually spread over the resonant island. The emergence of complex multi-peak distributions during phase mixing of trapped particle populations has been established in multiple physical systems[18,27]. Most of the energy range of zebra stripes, however, corresponds to non-resonant particles whose phase space trajectories are outside the separatrix. Formation of regular phase-space patterns by non-resonant particles is attributed to the periodic character of their azimuthal drift around Earth, which makes them effectively trapped. On the course of their azimuthal drift motion, non-resonant particles flow around the resonance island repeatedly at frequencies that depend on their proximity to the separatrix, which causes stretching and folding of their distributions, leading to the formation of stripes.

The process is schematically illustrated by Extended Data Fig. 2. Consider evolution of a group of non-resonant particles distributed in azimuthal angle but with the same energy and radial position taken from a smooth distribution function that initially is independent of the azimuthal angle (panels **a**, **c**). All particles of the group therefore have the same value of phase space density. Particle motion around the resonant island will stretch the distribution and eventually create a fold. Panel **d** corresponds to the moment when points A, B and C are aligned along the azimuthal angle where a spacecraft observation is taken. Since the phase space density is conserved along particle trajectories, points A, B and C of the measured energy spectrum correspond to the same value of the phase space density. The energy spectrum among these three points, however, is produced by particles with different initial energies and phase space density values. Consequently, the A, B, C fold in the particle distribution results in a local peak in the measured profile of the electron phase space density (panel **b**). Multiple folds will create multiple local peaks in the phase space density, which constitute zebra patterns observed by the spacecraft.

As can be seen from Supplementary Video 1, with the model parameters chosen for quiet-time geomagnetic conditions it takes about 1 day for a distinct zebra pattern to emerge from an initially smooth electron distribution. Long-term evolution of a zebra pattern under steady-state conditions is illustrated in Extended Data Fig. 3, which shows electron intensity ($j = p^2 f$) as a function of energy along the surface of a constant first adiabatic invariant from the simulation illustrated in Fig. 3 and Supplementary Video 2. Stretching and folding leads to a continuous increase of the number of stripes associated with the decrease in their width and spacing. This can explain variations in the zebra patterns observed by RBSPICE A (top panels of Fig. 1).

If the stretching and folding process could continue uninterrupted, a zebra pattern monitored with a finite energy resolution detector would eventually merge into a seemingly continuous spectrum. However, in the 6-month interval between March and August 2013 zebra patterns were observed at all perigee passes. This can be explained by a disruption of the stretching and folding process by additional perturbations of the co-rotation region which reset energetic electron distributions to new initial conditions. This explanation is supported by the fact that the zebra patterns are much more variable during intervals of enhanced geomagnetic activity, when the amplitude of the solar-wind-driven perturbations penetrating inside the co-rotation region increases.

To examine how additional disturbances can modify the zebra patterns, we considered one possible source of such disturbances: magnetospheric substorms, which produce global perturbations of the fields that can penetrate deep into the inner magnetosphere. The following electric field term was added to the right-hand side of the first equation in system (1):

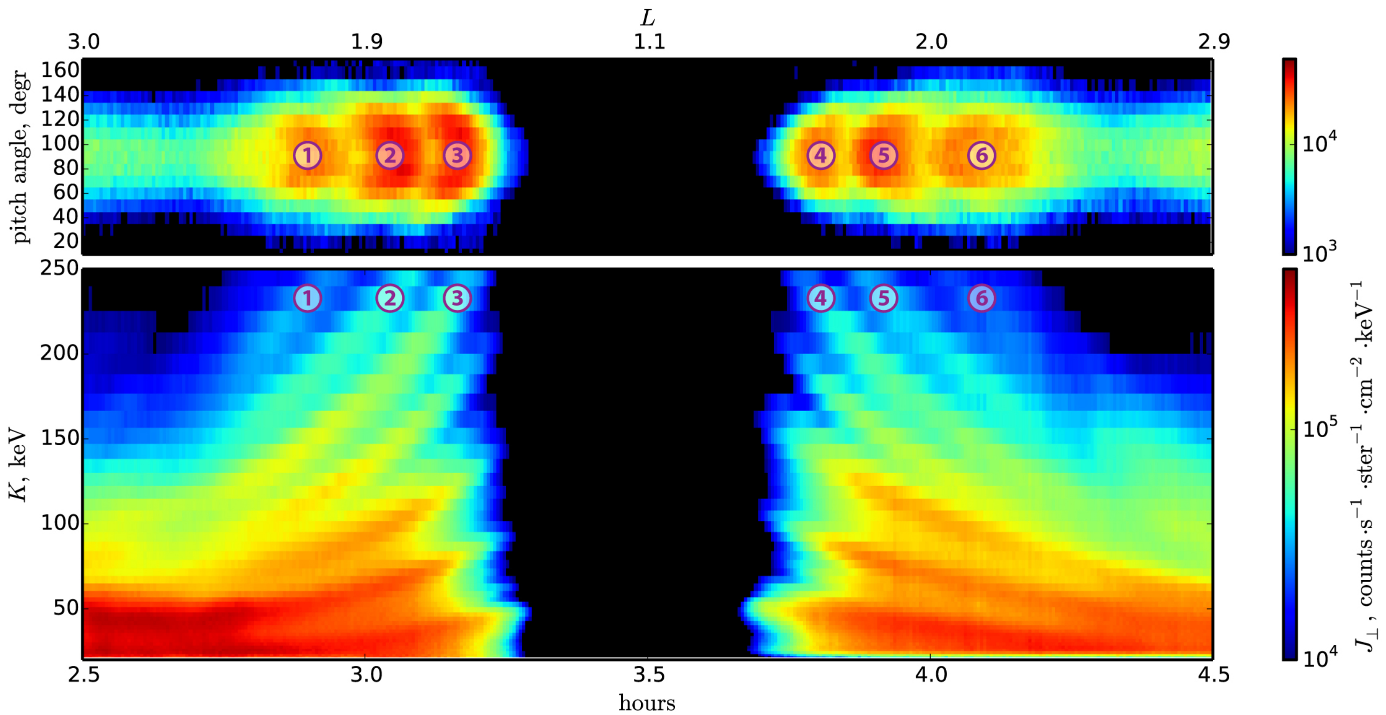
$$E_{\varphi 1} = E_1 \sum_{i=1}^{5} \sin(\varphi - \omega_i t + \psi_i) \qquad (3)$$

where $\omega_i = 2\pi/T_i$, and T_i were evenly distributed between 2 and 4 h corresponding to the statistically determined range of the substorm recurrence, and the phase shifts ψ_i were randomly selected between 0 and 2π. According to statistical analysis, the inductive electric fields associated with substorm dipolarizations are of the order of $5\ \mathrm{mV\,m^{-1}}$ at $L = 4.2$ (ref. 28). Although there are no direct measurements of substorm associated electric field at the inner-belt altitudes, it can be expected that the perturbations have become much weaker by the time they reach the inner belt. The amplitude $E_1 = 0.32\ \mathrm{mV\,m^{-1}}$ was therefore chosen such that the root-mean-square amplitude of the perturbation is $0.5\ \mathrm{mV\,m^{-1}}$, that is, an order of magnitude less than the characteristic amplitudes observed at higher L-shells. The model results (Fig. 1h and Supplementary Video 3) are qualitatively similar to energetic electron distributions observed during geomagnetic storms.

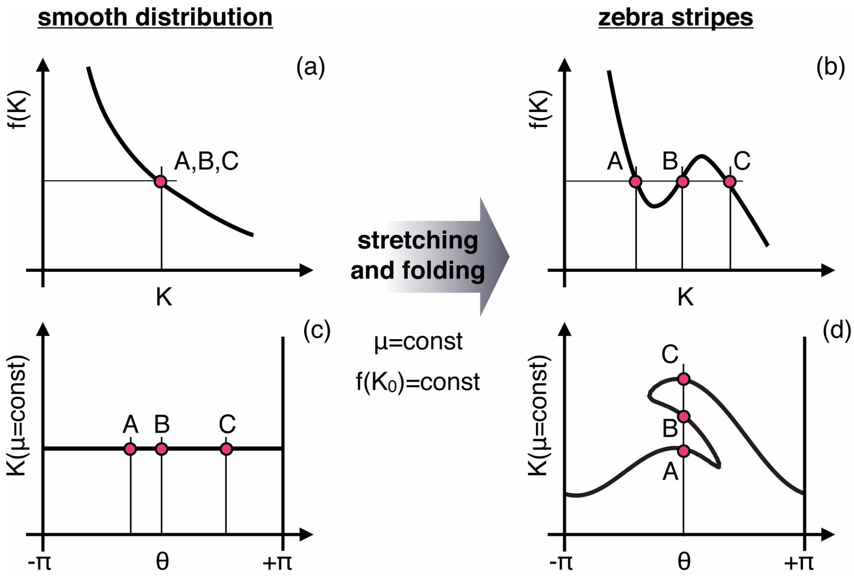
Supplementary Video 4 shows the results from a simulation with a slightly modified perturbation electric field, where the upper limit of the substorm periodicity in equation (3) was increased from 4 to 6 h. Considerable differences in zebra patterns between this and the previous simulations show that the structure of the zebra patterns is very sensitive to details of perturbations in the field, which is in agreement

with large diversity of zebra patterns observed during disturbed geomagnetic conditions (Fig. 1e–g).

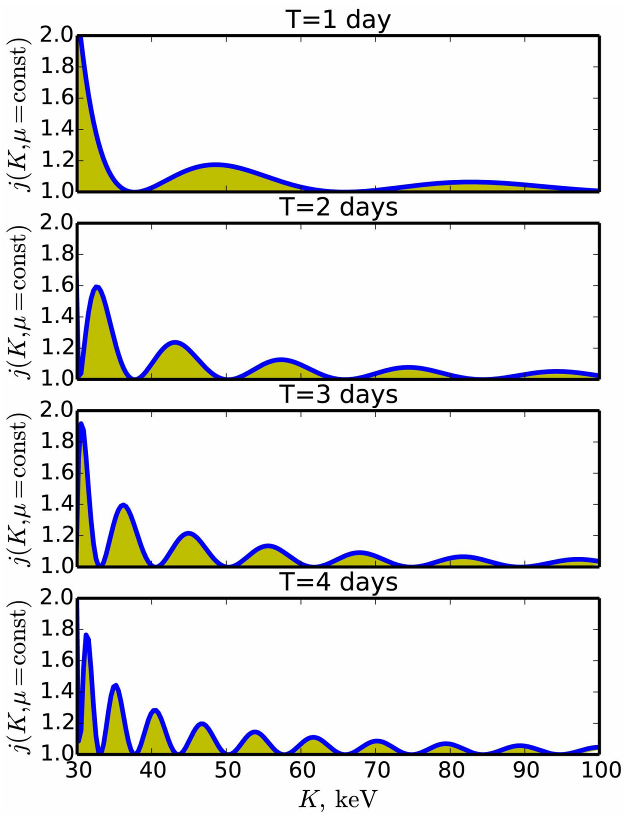
24. Northrop, T. G. & Teller, E. Stability of the adiabatic motion of charged particles in the Earth's field. *Phys. Rev.* **117,** 215–225 (1960).
25. Northrop, T. G. *The Adiabatic Motion of Charged Particles* (Interscience, 1963).
26. Cary, J. R. & Brizard, A. J. Hamiltonian theory of guiding-center motion. *Rev. Mod. Phys.* **81,** 693–738 (2009).
27. Birdsall, C. K. & Langdon, A. B. in *Plasma Physics via Computer Simulation* (eds Cowley, S., Stott, P. & Wilhelmsson, H.) Ch. 5–15 (Plasma Physics Series, Institute of Physics Publishing, 1991).
28. Ohtani, S. *et al.* Inductive electric fields in the inner magnetosphere during geomagnetically active periods. *J. Geophys. Res.* **115,** A00I14 (2010).



Extended Data Figure 1 | A zebra stripe pattern from the RBSPICE A perigee pass on 30 March 2013. Differential intensity (shown with colour) of energetic electrons as function of time along a perigee pass through the inner radiation belt (radial distance L in Earth radii is marked on the top). Bottom panel, the energy dependence (in keV) of differential intensity for electrons with approximately 90° pitch angles. Upper panel, the pitch-angle distribution of the intensity of 230 keV electrons. It shows highly pronounced stripes that are curved towards more field-aligned particles. The ringed numbers plotted along the 230 keV line in the bottom panel and along the 90° values in the upper panel show the correspondence of stripes in the energy and pitch-angle spectra.



Extended Data Figure 2 | Formation of zebra stripes in the process of stretching and folding of the electron phase space density distribution. **a**, **c**, Smooth initial phase space density independent of the drift angle $f(K)$ (**a**); points A, B and C with different drift angle (θ) values but the same energy K (radial position) correspond to the same value of the phase space density (**c**). **b**, **d**, The moment in the stretching and folding process when points A, B and C are aligned along the azimuthal angle where a spacecraft observation is taken (**d**). Points A, B and C are separated in energy at the same level of the phase space density, which requires formation of a local peak in phase space density (**b**) corresponding to a stripe of the emerging zebra pattern.



Extended Data Figure 3 | Long-term evolution of a zebra pattern under steady-state conditions. Four panels show electron intensity ($j = p^2 f$) as function of energy along the surface of a constant adiabatic invariant (same as Fig. 3 and Supplementary Video 2) at different moments (top to bottom, time $T =$ 1, 2, 3 and 4 days) of a long-term simulation of electron motion in the presence of global diurnal oscillations in the electric field. Stretching and folding leads to a continuous increase of the number of stripes associated with the decrease in their width and spacing.

LETTER

doi:10.1038/nature13078

A fully photonics-based coherent radar system

Paolo Ghelfi[1], Francesco Laghezza[1], Filippo Scotti[1], Giovanni Serafino[2], Amerigo Capria[3], Sergio Pinna[2], Daniel Onori[2], Claudio Porzi[2], Mirco Scaffardi[1], Antonio Malacarne[1], Valeria Vercesi[2], Emma Lazzeri[1], Fabrizio Berizzi[3,4] & Antonella Bogoni[1]

The next generation of radar (radio detection and ranging) systems needs to be based on software-defined radio to adapt to variable environments, with higher carrier frequencies for smaller antennas and broadened bandwidth for increased resolution[1–4]. Today's digital microwave components (synthesizers and analogue-to-digital converters) suffer from limited bandwidth with high noise at increasing frequencies[5–7], so that fully digital radar systems can work up to only a few gigahertz, and noisy analogue up- and downconversions are necessary for higher frequencies. In contrast, photonics provide high precision and ultrawide bandwidth[8,9], allowing both the flexible generation of extremely stable radio-frequency signals with arbitrary waveforms up to millimetre waves[10–22], and the detection of such signals and their precise direct digitization without downconversion[23–26]. Until now, the photonics-based generation and detection of radio-frequency signals have been studied separately and have not been tested in a radar system. Here we present the development and the field trial results of a fully photonics-based coherent radar demonstrator carried out within the project PHODIR[27]. The proposed architecture exploits a single pulsed laser for generating tunable radar signals and receiving their echoes, avoiding radio-frequency up- and downconversion and guaranteeing both the software-defined approach and high resolution. Its performance exceeds state-of-the-art electronics at carrier frequencies above two gigahertz, and the detection of non-cooperating aeroplanes confirms the effectiveness and expected precision of the system.

The performance of the transceiver we developed is summarized in Table 1, together with a picture of the project demonstrator (Fig. 1). The transceiver enables radar systems with unprecedented frequency flexibility, signal quality and receiver resolution at high carrier frequency. The radar demonstrator has also been tested in a real environment.

The photonic generation of radio-frequency (RF) signals, avoiding microwave oscillators, can be realized either by means of opto-electronic oscillators[21,22] or by the heterodyning of two lasers in a photodiode[10]. The opto-electronic oscillators show excellent stability up to 40 GHz with a miniaturized footprint, but their tunability is limited to a few gigahertz, whereas the heterodyning can generate any frequency up to the photodiode bandwidth, and is therefore more promising. Nevertheless, it requires the two lasers to be phase-locked. This can easily be obtained by selecting any pair of modes from the spectrum of mode-locked lasers (MLLs)[12,13], thanks to their intrinsic phase-locking condition. Moreover, selecting laser modes with variable wavelength detuning allows the flexible production of RF carriers with tunable frequency. Finally, the heterodyne method allows us to add an arbitrary modulation on the generated RF signal by modulating one of the beating laser modes.

The receivers for radar systems based on software-defined radio need analogue-to-digital converters (ADCs) that accept input signals over several tens of gigahertz, with fast sampling rates suitable for broadband signals and a high signal-to-noise ratio. Electronic ADCs are limited by the aperture jitter of the electronic sampling process—that is, the fluctuations of the interval duration between two consecutive samples—which spoils the accuracy of the digitized signal by adding a noise floor. Optical sampling can overcome this limitation[23]: a train of short optical pulses can be directly modulated by the RF signal to be digitized, and then the peak amplitude of each optical pulse can be detected and digitized. Thus, sampling rates of several gigahertz are achievable, and the timing jitter of the sampling pulses can be much less than 100 fs. MLLs are particularly suitable pulse sources, with narrow pulsewidths and low temporal jitter.

The basic idea of the PHODIR project ('Photonics-based fully digital radar') is to use a single MLL for the realization of both the RF generation and sampling, obtaining an entire optics-based transceiver, as sketched in Fig. 2 (left), reducing the impact of photonics on the system costs. In the band-pass sampling process, which requires the sampling frequency to be at least twice the bandwidth of the pass-band signal, the carrier frequency of the sampled signal must not be a multiple of the sampling frequency. Unfortunately, this is the case when the same MLL is used for both generating and sampling the radar signal. This problem can be fixed by shifting one of the two MLL modes before heterodyning. The frequency shift must be driven by an RF oscillator with a negligible phase noise compared to the MLL, which is true if low-frequency oscillators

Table 1 | Performance of the photonics-based radar transceiver

Parameter	Photonics-based transceiver	State-of-the-art electronics transceiver
	Transmitter	
Carrier frequency	Flexible direct generation up to 40 GHz	Direct generation below 2 GHz; up-conversions above 2 GHz
Signal jitter	<15 fs; integrated from 10 kHz to 10 MHz	Typical >20 fs, integrated from 10 kHz to 10 MHz
Signal-to-noise ratio	>73 dB MHz^{-1}	>80 dB MHz^{-1}
Spurious-free dynamic range	>70 dBc	>70 dBc
Instantaneous bandwidth	200 MHz, easily extendable with MLL at higher repetition rate	<2 GHz
	Receiver	
Input carrier frequency	Up to 40 GHz, with direct RF undersampling	<2 GHz; down-conversions at higher frequencies
Instantaneous bandwidth	200 MHz, easily extendable with MLL at higher repetition rate	<2 GHz
Sampling jitter	<10 fs, integrated from 10 kHz to 10 MHz	Typical >100 fs, integrated from 10 kHz to 10 MHz
Spurious-free dynamic range	50 dB	>70 dB
Effective number of bits	>7 for carrier frequency up to 40 GHz	<8 for carrier frequency <2 GHz

Test results of the photonics-based transceiver are shown versus state-of-the-art electronic radar transceivers.

[1]National Laboratory of Photonic Networks, Inter-university National Consortium for Telecommunications (CNIT), Via Moruzzi 1, 56124 Pisa, Italy. [2]TeCIP Institute, Scuola Superiore Sant'Anna, Via Moruzzi 1, 56124 Pisa, Italy. [3]National Laboratory of Radars and Surveillance Systems, Inter-university National Consortium for Telecommunications (CNIT), Galleria Gerace 18, 56126 Pisa, Italy. [4]Department of Information Engineering, University of Pisa, Via Caruso 16, 56122 Pisa, Italy.

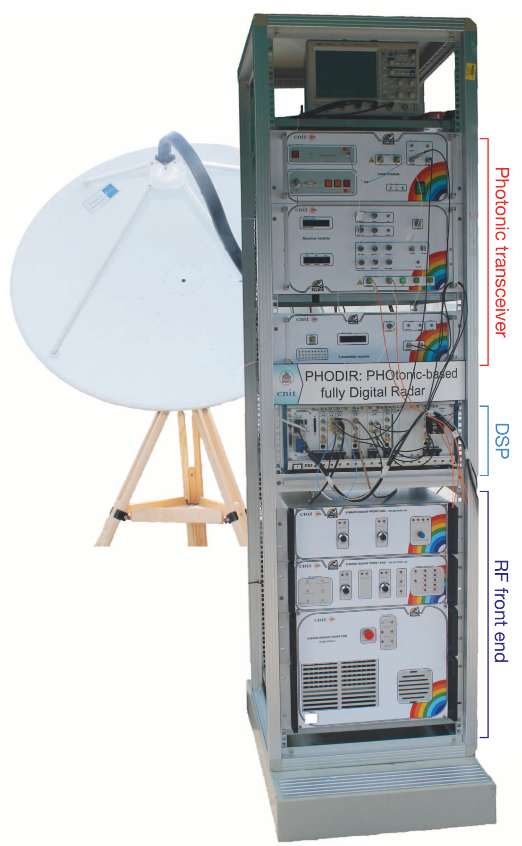


Figure 1 | The photonics-based radar transceiver. A picture (taken by A.B.) of the field-trial demonstrator. DSP, digital signal processor.

are used[28]. Figure 2 (right) shows details of the PHODIR transceiver. In the photonics-based RF generator, two tunable optical filters select two modes from the spectrum of the MLL. One is modulated by the baseband radar pulse in a Mach–Zehnder modulator (MZM), whereas the other is shifted in frequency. In the photonics-based ADC, the optical pulses from the MLL are modulated in a MZM driven by the received echo signal. The modulated pulses are then parallelized in a serial-to-parallel converter into multiple lower-rate sample streams so that, after opto-electronic conversion, multiple electronic ADCs with low sampling frequency and high precision can be used to digitize the pulse peak power carrying the sample information. The parallel digital samples are then interleaved in the digital signal processor to reconstruct the original signal. To match the pulse width to the photodiode bandwidth, spools of dispersion compensating fibre are inserted to broaden the optical pulses.

The precision of photonics is particularly valuable in coherent radar systems, which derive the speed of the targets from the Doppler frequency shift obtained by comparing the instantaneous frequency of the echoed signal with that of a continuous-wave reference signal. To allow this functionality, the PHODIR transceiver also generates the reference signal (see Fig. 2), exploiting a similar structure used for the radar pulse but without the modulation. Orthogonal sampling is implemented to detect the in-phase and quadrature components of the reference signal, allowing us to sample the reference signal only once and to derive the quadrature samples by shifting the in-phase samples[1].

Figures 3 and 4 show the results obtained when testing the transceiver in the laboratory. The MLL we used generates subpicosecond pulses with a repetition rate of 400 MHz. The transmitter that we used differs slightly from the scheme in Fig. 2 in that the laser modes are not selected by tunable filters. Instead, the spectrum of the MLL is modulated by the radar pulse at an intermediate frequency equal to the frequency shift, and then photodetected. At the photodiode there are several pairs of modes that beat together at different frequencies, including the desired one. Given that all the beatings at the same frequency have the same phase, at the photodiode output a clear signal can be selected by a RF filter at the desired frequency. Therefore, here the tunability is moved

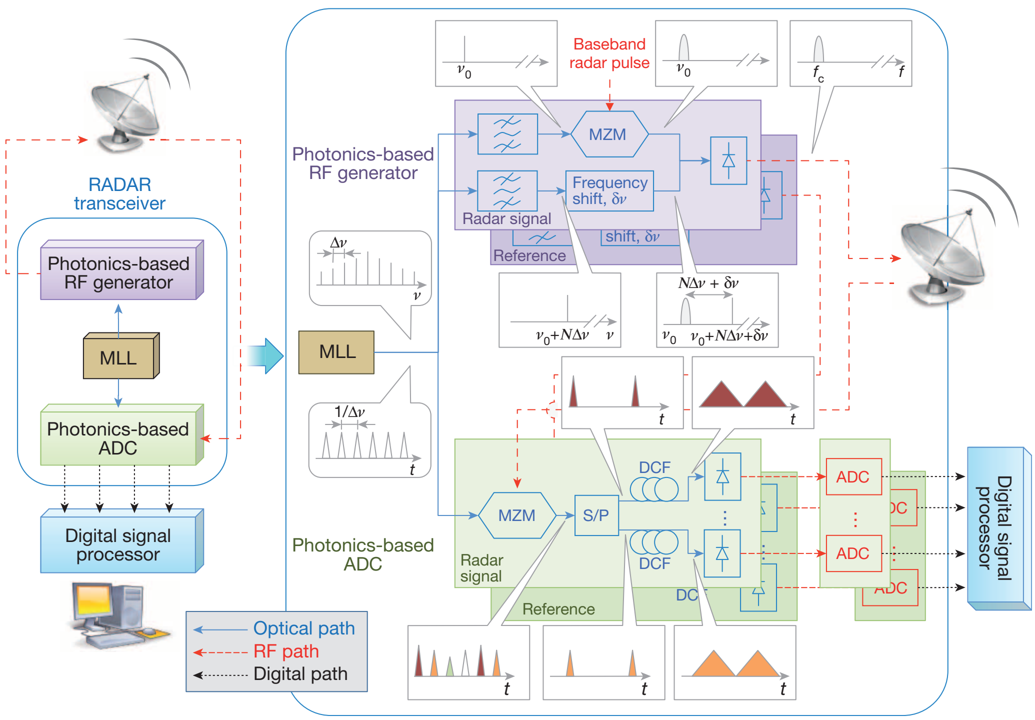


Figure 2 | Architecture of the photonics-based radar. Basic structure and detailed scheme of principle of the photonics-based coherent radar. S/P, series-to-parallel converter, DCF, dispersion compensating fibre. t, time; ν, optical frequency; $\Delta\nu$, MLL repetition frequency; $\delta\nu$, applied frequency shift; f, RF frequency; N, positive integer.

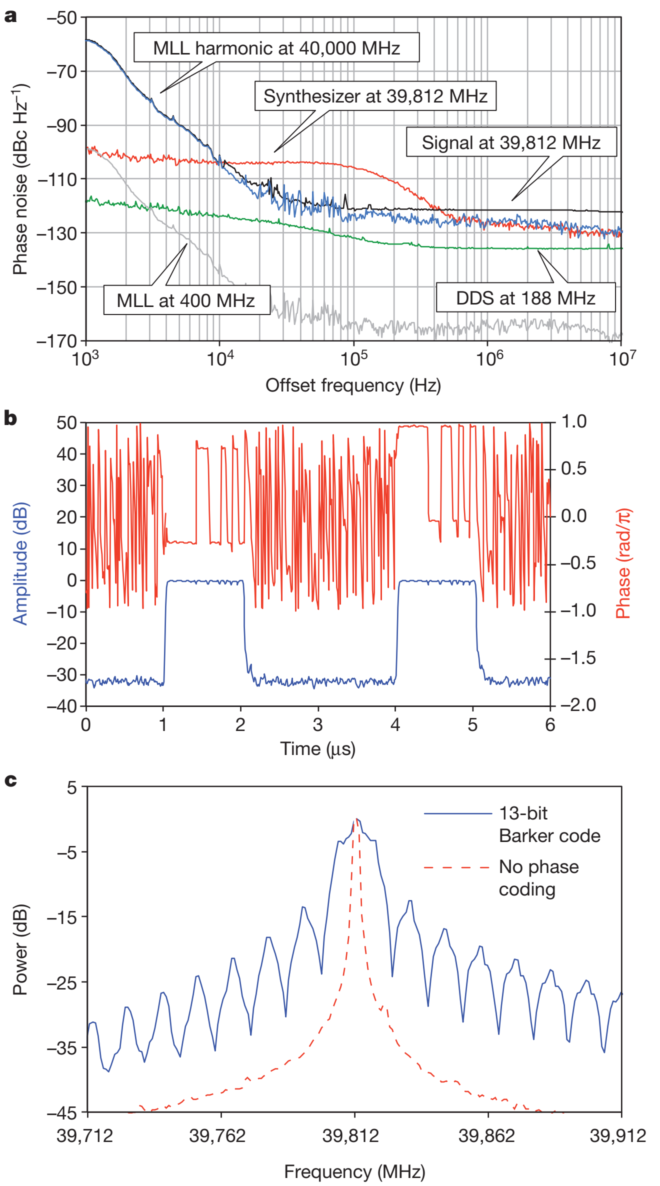


Figure 3 | Test results of the photonics-based RF generator. **a**, Phase noise curve of the RF signal optically generated at 39,812 MHz, compared with the phase noise curves of the fundamental frequency of the laser at 400 MHz, of its hundredth multiple at 40,000 MHz, of the continuous-wave intermediate frequency at 188 MHz generated by a direct digital synthesizer, and of a signal at 39,812 MHz generated by a synthesizer. **b**, Amplitude and phase transients of a 1-μs radar pulse phase-modulated by a 13-bit Barker code with carrier frequency of 39,812 MHz. **c**, RF spectrum of the 13-bit Barker coded radar pulse at 39,812 MHz, compared with the RF spectrum of the uncoded pulse.

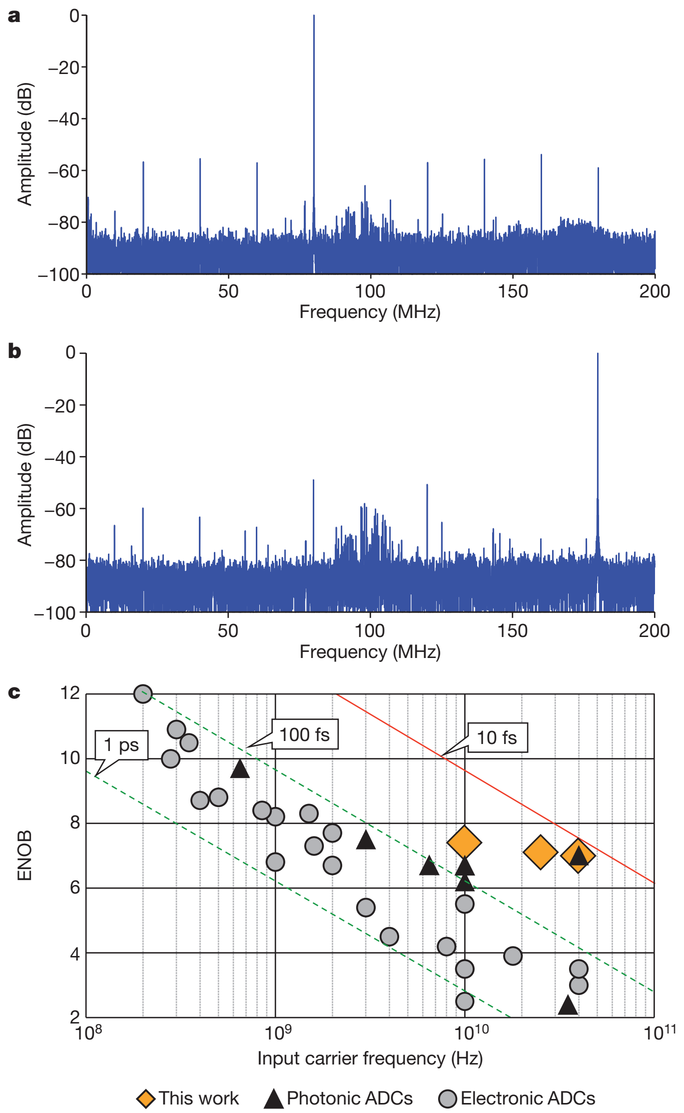


Figure 4 | Test results of the photonics-based ADC. **a**, **b**, Fast Fourier transform spectrum of the RF signal at 9,920 MHz (**a**) and 39,820 MHz (**b**), optically sampled at 400 × 10^6 samples per second and digitized at 4 × 100 × 10^6 samples per second. **c**, Extended Walden plot, comparing the performance of electronic and photonic ADCs.

in the RF domain by means of selectable RF filters. In any case, this does not affect the photonics-based approach because the stability and flexibility of the generated RF signals are only obtainable using photonics.

We tested the photonics-based RF generator, producing signals with carriers up to 40 GHz (limited by the photodiode bandwidth). As an example, Fig. 3 reports the results obtained at 39,812 MHz. Figure 3a reports the stability, in terms of the phase noise curve, of the optically generated RF carrier, which is obtained applying a shift of 188 MHz to the hundredth multiple of the MLL fundamental frequency. The graph also reports the phase noise of the detected fundamental frequency of the laser at 400 MHz, of its hundredth multiple at 40,000 MHz, and of the continuous-wave intermediate frequency at 188 MHz generated by a direct digital synthesizer. As expected[12], the phase noise curve of the 40,000 MHz signal is equivalent to the fundamental phase noise curve of the laser shifted up by 40 dB. The signal at intermediate frequency shows significantly lower phase noise than that at 40,000 MHz, so the signal at 39,812 MHz substantially maintains the noise characteristics of the non-shifted signal. The higher noise floor at high offset frequency is induced by the necessary RF amplifiers, and does not substantially affect the signal stability. Given that the typical radar pulse repetition time is about 100 μs, the phase noise curve is of interest for offset frequencies from 10 kHz and up. The timing jitter can be calculated by integrating the phase noise over this range, which turns out to be 10 fs for the fundamental 400 MHz and its multiple at 40,000 MHz, and 15 fs for the shifted signal. For comparison, Fig. 3a also reports the phase noise curve of a signal at 39,812 MHz generated by a state-of-the-art synthesizer, which shows a jitter of 23 fs. Our scheme therefore performs significantly better than the synthesizer. A spectrum analysis of the photonics-generated carrier also shows a signal-to-noise ratio higher than 73 dB on a noise bandwidth of 1 MHz, and a spurious-free dynamic range of >70 dBc (decibels relative to the carrier power). Analogous results have been obtained generating signals in the range from a few hertz to 40 GHz by substituting the RF filter after the photodiode with others at a different central frequency, and setting the direct digital

synthesizer to generate a different intermediate frequency. This demonstrates that the proposed method can effectively generate very stable RF signals at arbitrary frequencies over an unprecedented broad range. Performance improvements can be obtained by exploiting photonic integration techniques.

By setting the direct digital synthesizer to generate modulated signals at intermediate frequency rather than continuous-wave signals, we can move the modulation signals on the desired RF carrier. Radar pulses can therefore be directly produced on the RF carrier by digitally generating pulse-modulated signals at intermediate frequency. Moreover, by adding a phase modulation to the digital signal from the synthesizer, phase-coded radar pulses can be realized for increasing the radar resolution through spectral broadening. Figure 3b reports the amplitude and phase transients of the RF signal at 39,812 MHz that are produced when the intermediate frequency is modulated by 1-μs rectangular pulses with the typical 13-bit Barker code phase modulation. Figure 3c clearly shows the expected pulse spectrum broadening from 1 MHz to 13 MHz, induced by the phase modulation. The bandwidth could be extended up to half the repetition rate of the MLL (200 MHz). These results confirm the capability of the scheme to generate arbitrary amplitude-modulated and/or phase-modulated signals[19].

Figure 4 reports the results obtained by testing the photonics-based ADC of the PHODIR system[26]. The RF signal is optically sampled at 400×10^6 samples per second in a MZM with 40-GHz bandwidth. The modulated optical pulses are parallelized into four streams at 100×10^6 samples per second, exploiting an integrated 1×4 $LiNbO_3$ switching matrix, ensuring an extinction ratio—that is, a power ratio between a selected output port and the three disabled ones—of >25 dB (ref. 29). The pulses are broadened in four spools of dispersion compensating fibre and digitized by four electronic ADCs synchronized with the pulse peaks. Finally, a digital signal processor interleaves the samples. The photonics-based ADC is tested with input continuous-wave signals up to 40 GHz. As examples, Fig. 4a and b report the calculated fast Fourier transform for signals at 9,920 MHz and 39,820 MHz, both with $4 \times 10{,}000$ points, after the digital processing. The undersampling process aliases the signals in the Nyquist bandwidth (direct current to 200 MHz), at 80 MHz and 180 MHz, respectively. The noise floor in the two graphs appears at different levels owing to the increasing influence of the sampling jitter for increasing signal frequency. The residual spurious tones after the data post-processing are several tens of decibels below the signal, enabling a spurious-free dynamic range of >50 dB. A higher spurious-free dynamic range could be obtained with a more precise sample parallelization process. By integrating the fast Fourier transform it is possible to calculate the signal-to-noise and distortion ratios, and from this the effective number of bits. In Fig. 4c the so-called 'Walden plot' is shown, reporting the performance comparison in terms of input carrier frequency and the effective number of bits of the best electrical and photonic ADCs reported so far[7], and of the PHODIR receiver. The theoretical upper bounds caused by different sampling jitters are also reported. The jitter of the MLL obtained by integrating its phase noise curve from 10 kHz is about 10 fs, whereas the best electronic ADCs show an aperture jitter of about 100 fs. As a consequence, our system reaches an effective number of bits of 7.4 for an input signal at 9,920 MHz, and 7 at 39,820 MHz, performing significantly better than the reported electrical ADCs. At 40 GHz, the system performance is close to the theoretical upper bound for 10 fs of aperture jitter, but it moves away from the upper bound at lower signal frequencies, indicating that at lower frequency the system's effective number of bits is limited by the spurious tones, whereas at higher frequency it is mostly limited by the time jitter of the MLL, because the spurious tones become comparable with the noise floor. Improvements in the effective number of bits could be achieved by including a differential detection scheme to suppress the detected noise[24], and an improved switching matrix with higher extinction to reduce the power of the spurious tones. Moreover, a larger instantaneous bandwidth could be achieved with higher sampling rates.

The test results of the PHODIR transceiver are summarized in Table 1, and compared with the performance of state-of-the-art electronic radar transceivers[6]. The advantages of the photonic approach are evident in the extreme frequency flexibility over tens of gigahertz, in the arbitrary modulation capability, and in the precision of the digitization for any input frequency. We believe that these features will enable the use of software-defined radio in future radar systems.

We also tested the PHODIR transceiver reported above in a field-trial demonstrator to determine its effectiveness in a radar system. For this purpose, we used a RF front-end (RF circulator, switches, amplifiers, filters and bistatic antenna) for a signal carrier at 9,900 MHz. Non-cooperating (that is, normally operating and unaware of our detection) aeroplanes were detected as dynamic targets. For example, Fig. 5a reports the trajectory of an observed aeroplane during takeoff. The range–velocity map of the target at point A of the trajectory is reported in Fig. 5b. The velocity and position of the target were cross-checked by trajectory data provided by public websites (http://www.flightradar24.com/). At the detection point, the known radial velocity of the target is about 96 km h^{-1} (for a real speed of about 400 km h^{-1}), and the known distance is about 5.5 km. The corresponding detected signal shows the target at 5.4 km with a radial velocity of 95 km h^{-1}. Figure 5c shows a

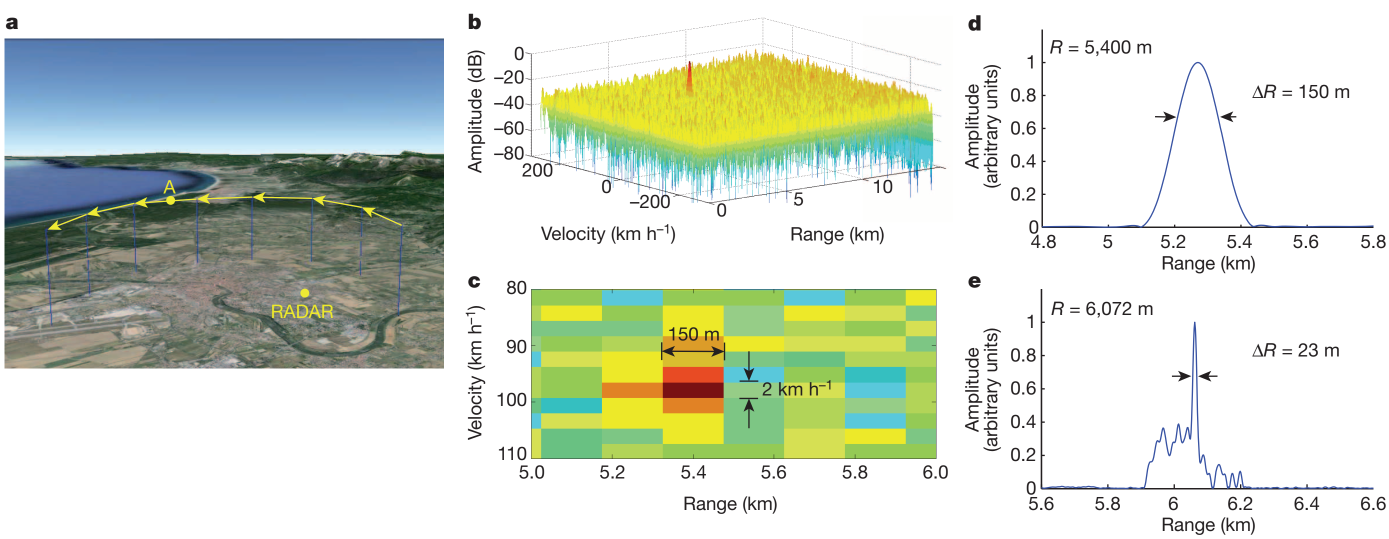


Figure 5 | Test results of the field-trial demonstrator. **a**, Trajectory of a target aeroplane. **b**, Range–velocity map at point A in the trajectory. **c**, Zoom showing the range and velocity resolution. **d**, **e**, Range profile without (**d**) or with (**e**) pulse coding.

two-dimensional zoom of the range–velocity map, with distance resolution of 150 m and velocity resolution of 2 km h^{-1}. The target shows a detected echo 45 dB above the background. Finally, a 13-bit Barker code was modulated on the phase of the radar pulses to increase the resolution. Figure 5d and e report the range profile collected with and without this pulse coding, showing an increase in precision from 150 m to 23 m.

The PHODIR project has pioneered a fully photonics-based radar system, enabling the software-defined radio approach with a performance exceeding that of state-of-the-art radar systems. In particular, today's electronic transceivers cannot achieve the same frequency range without the use of several parallel architectures, and do not provide an equivalent precision, especially at high carrier frequencies. A field-trial demonstrator has proved the effectiveness and the expected precision of the photonic architecture. This proof-of-concept system has been realized using commercial devices, except for the switching matrix which was developed within the EU programme NEXPRESSO[29]. We are now planning to develop dedicated photonic integrated circuits, which could fully realize the potential of the photonic approach, leading to compact, flexible systems with high performance for the most advanced applications, such as simultaneous surveillance and communications. The use of photonics in radar systems also allow additional functionalities, which make its use even more convenient. For example, tunable time delay can be applied to the lasers transporting the RF signals to control the beamforming at arrayed antennas[30]. Moreover, photonics permits the use of optical fibres for transporting the radar signals to and from the antenna, allowing its remotization with negligible losses. Therefore, the architecture we propose, as sketched in Fig. 2, should open up new frontiers in the field of radar, enabling future smart multifunction surveillance systems.

METHODS SUMMARY

We use the OneFive Origami passive MLL with a pulse repetition frequency of 400 MHz and an optical bandwidth of about 8 nm at a wavelength of 1,560 nm. In photonics-based signal generation, the frequency shift of one of the modes of the MLL can be realized, for example, by single-sideband modulation with carrier suppression. The state-of-the-art synthesizer used as a reference in the analysis of the phase noise of the photonics-generated signal is the Agilent E8257D. The integrated 1×4 $LiNbO_3$ switching matrix employed at the receiver stage consists of three dual-output MZMs in a cascaded configuration. The chromatic dispersion of the spools of dispersion compensating fibre before the photodiodes is -300 ps nm^{-1}. The electronic ADCs are the data acquisition boards NI PXIe5122 from National Instruments, with 14-bit resolution and 100×10^6 samples per second. The digital processing implemented to interleave the samples also suppresses the spurious tones caused by the nonlinearity of the optical modulator and the different attenuations and delays of the four parallel branches of the photonics-based receiver. The synthesizer used for testing the photonic ADC is the Agilent E8257D. In the RF front-end of the PHODIR demonstrator, the RF filter at the receiver has a bandwidth of 40 MHz. Therefore both the received echo and the reference signal are optically sampled at 400×10^6 samples per second, and digitized by acquiring their aliased replica with ADCs at 80×10^6 samples per second. The non-modulated 1-μs-long radar pulses give a range resolution of 150 m, with a velocity resolution of 2 km h^{-1} given by the coherent integration time of 20 ms. In the field trial, we ran a preliminary system calibration focusing on static targets, detecting objects in a range up to 30 km with a transmitted RF peak power of about 20 W.

Received 14 October 2013; accepted 23 January 2014.

1. Skolnik, M. L. *Introduction to Radar Systems* 3rd edn (McGraw-Hill, 1980).
2. Haykin, S. Cognitive radar: a way of the future. *IEEE Signal Process. Mag.* **23,** 30–40 (2006).
3. Ravenni, V. Performance evaluations of frequency diversity radar system. *Proc. Eur. Microwave Conf.* 1715–1718, http://dx.doi.org/10.1109/eumc.2007.4405545 (2007).
4. Tsui, J. B. *Digital Techniques for Wideband Receivers* 2nd edn (SciTech, 2004).
5. Scheer, J. A. & Kurtz, J. L. *Coherent Radar Performance Estimation* (Artech House, 1993).
6. Richards, M. A., Scheer, J. A. & Holm, W. A. *Principle of Modern Radar: Basic Principles* (SciTech Publishing, 2010).
7. Walden, R. *Analog-to-Digital Conversion in the Early Twenty-First Century* (Wiley Encyclopedia of Computer Science and Engineering, 2008).
8. Capmany, J. & Novak, D. Microwave photonics combines two worlds. *Nature Photon.* **1,** 319–330 (2007).
9. Yao, J. Microwave photonics. *J. Lightwave Technol.* **27,** 314–335 (2009).
10. Goldberg, L., Esman, R. D. & Williams, K. J. Generation and control of microwave signals by optical techniques. *IEEE Proc. J.* **139,** 288–295 (1992).
11. Khan, M. H. *et al.* Ultrabroad-bandwidth arbitrary radiofrequency waveform generation with a silicon photonic chip-based spectral shaper. *Nature Photon.* **4,** 117–122 (2010).
12. Serafino, G. *et al.* Stable optically generated RF signals from a fibre mode-locked laser. *Proc. 23rd IEEE Photonics Soc. Ann. Meet.* Abstr. TuK4, 193–194, http://dx.doi.org/10.1109/photonics.2010.5698824 (2010).
13. Yilmaz, T., DePriest, C. M., Turpin, T., Abeles, J. H. & Delfyett, P. J. Toward a photonic arbitrary waveform generator using a modelocked external cavity semiconductor laser. *IEEE Photonics Technol. Lett.* **14,** 1608–1610 (2002).
14. Chou, J., Han, Y. & Jalali, B. Adaptive RF-photonic arbitrary waveform generator. *IEEE Photonics Technol. Lett.* **15,** 581–583 (2003).
15. McKinney, J. D., Leaird, D. E. & Weiner, A. M. Millimeter-wave arbitrary waveform generation with a direct space-to-time pulse shaper. *Opt. Lett.* **27,** 1345–1347 (2002).
16. Lin, I. S., McKinney, J. D. & Weiner, A. M. Photonic synthesis of broadband microwave arbitrary waveform applicable to ultrawideband communication. *IEEE Microwave Wireless Components Lett.* **15,** 226–228 (2005).
17. Chi, H. & Yao, J. P. An approach to photonic generation of high frequency phase-coded RF pulses. *IEEE Photonics Technol. Lett.* **19,** 768–770 (2007).
18. Li, Z., Li, W., Chi, H., Zhang, X. & Yao, J. Photonic generation of phase-coded microwave signal with large frequency tunability. *IEEE Photonics Technol. Lett.* **23,** 712–714 (2011).
19. Ghelfi, P., Scotti, F., Laghezza, F. & Bogoni, A. Phase coding of RF pulses in photonics-aided frequency-agile coherent radar systems. *IEEE J. Quantum Electron.* **48,** 1151–1157 (2012).
20. Ghelfi, P., Scotti, F., Laghezza, F. & Bogoni, A. Photonic generation of phase-modulated RF signals for pulse compression techniques in coherent radars. *J. Lightwave Technol.* **30,** 1638–1644 (2012).
21. Maleki, L. *et al.* High performance, miniature hyper-parametric microwave photonic oscillator. *Proc. IEEE Freq. Control Symp.* 558–563, http://dx.doi.org/10.1109/freq.2010.5556265 (2010).
22. Maleki, L. The optoelectronic oscillator. *Nature Photon.* **5,** 728–730 (2011).
23. Valley, G. C. Photonic analog-to-digital converters. *Opt. Exp.* **15,** 1955–1982 (2007).
24. Khilo, A. *et al.* Photonic ADC: overcoming the bottleneck of electronic jitter. *Opt. Exp.* **20,** 4454–4469 (2012).
25. Chou, J., Conway, J. A., Sefler, G. A., Valley, G. C. & Jalali, B. Photonic bandwidth compression front end for digital oscilloscopes. *J. Lightwave Technol.* **27,** 5073–5077 (2009).
26. Laghezza, F. *et al.* Jitter-limited photonic analog-to-digital converter with 7 effective bits for wideband radar applications. *Proc. IEEE Radar Conf.* 1–5, http://dx.doi.org/10.1109/radar.2013.6586075 (2013).
27. *Photonic-based Fully Digital Radar* (PHODIR) http://www.phodir.eu (2009).
28. Ghelfi, P., Scotti, F., Nguyen, A. T., Serafino, G. & Bogoni, A. Novel architecture for a photonics-assisted radar transceiver based on a single mode-locking laser. *IEEE Photonics Technol. Lett.* **23,** 639–641 (2011).
29. Pierno, L. *et al.* Optical switching matrix as time domain demultiplexer in photonic ADC. *Proc. Eur. Microwave Integr. Circuits Conf.* Abstr. EuMIC03.3, INSPEC accession number 13990947 (2013).
30. Ghelfi, P. *et al.* Photonic generation and independent steering of multiple RF signals for software defined radars. *Opt. Exp.* **21,** 22905–22910 (2013).

Acknowledgements This work was supported by the ERC projects PHODIR (contract number 239640) and PREPARE (contract number 324629), and by the EU NEXPRESSO programme through the project INSIDE with Selex Sistemi Integrati S.p.A. (now Selex ES S.p.A.).

Author Contributions A.B. coordinated all the activities of the PHODIR project. A.B. and P.G. designed the architecture of the photonics-based transceiver and wrote the paper. F.L., A.C. and F.B. defined the radar parameters and designed the RF front end. F.L., F.S., S.P., G.S., E.L. and D.O. implemented the photonic subsystems. S.P. designed and developed the electronic controls of the machine that separates the samples into parallel streams and of the front panels of the radar demonstrator. F.S. assembled the demonstrator. F.L. implemented the digital processing tools. P.G., G.S., M.S., E.L. and A.B. analysed and discussed the results from the photonics-based transmitter. P.G., F.L., F.S., S.P., D.O., A.M. and A.B. analysed and discussed the results from the photonics-based receiver. F.L., F.S., S.P., G.S. and D.O. collected and processed the data of the field trial. P.G., F.L., F.S., G.S., S.P., D.O. and A.B. analysed and discussed the results of the field trial. C.P., V.V., P.G. and A.B. discussed the possible development of the photonics-based transceiver with integrated photonics techniques. P.G., G.S., F.L., F.S. and A.B. commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.B. (antonella.bogoni@cnit.it).

LETTER

doi:10.1038/nature13030

Sulphide oxidation and carbonate dissolution as a source of CO_2 over geological timescales

Mark A. Torres[1], A. Joshua West[1] & Gaojun Li[2]

The observed stability of Earth's climate over millions of years is thought to depend on the rate of carbon dioxide (CO_2) release from the solid Earth being balanced by the rate of CO_2 consumption by silicate weathering[1]. During the Cenozoic era, spanning approximately the past 66 million years, the concurrent increases in the marine isotopic ratios of strontium, osmium and lithium[2–4] suggest that extensive uplift of mountain ranges may have stimulated CO_2 consumption by silicate weathering[5], but reconstructions of sea-floor spreading[6] do not indicate a corresponding increase in CO_2 inputs from volcanic degassing. The resulting imbalance would have depleted the atmosphere of all CO_2 within a few million years[7]. As a result, reconciling Cenozoic isotopic records with the need for mass balance in the long-term carbon cycle has been a major and unresolved challenge in geochemistry and Earth history. Here we show that enhanced sulphide oxidation coupled to carbonate dissolution can provide a transient source of CO_2 to Earth's atmosphere that is relevant over geological timescales. Like drawdown by means of silicate weathering, this source is probably enhanced by tectonic uplift, and so may have contributed to the relative stability of the partial pressure of atmospheric CO_2 during the Cenozoic. A variety of other hypotheses[8–10] have been put forward to explain the 'Cenozoic isotope-weathering paradox', and the evolution of the carbon cycle probably depended on multiple processes. However, an important role for sulphide oxidation coupled to carbonate dissolution is consistent with records of radiogenic isotopes[2,3], atmospheric CO_2 partial pressure[11,12] and the evolution of the Cenozoic sulphur cycle, and could be accounted for by geologically reasonable changes in the global dioxygen cycle, suggesting that this CO_2 source should be considered a potentially important but as yet generally unrecognized component of the long-term carbon cycle.

It has been proposed that excess CO_2 uptake from enhanced silicate weathering over the Cenozoic could be balanced by a net decrease in the size of the organic C reservoir[8]. However, recent work has shown that orogenic activity generally increases organic C burial and stabilizes petrogenic organic C, through graphitization[13]. Model results also suggest that global changes in organic matter burial during the Cenozoic were insufficient to balance excess CO_2 consumption (see, for example, refs 10, 14).

The release of CO_2 from the thermal decomposition of carbonate minerals during metamorphism has been suggested to be the missing source of CO_2 required to balance the Cenozoic C cycle[9]. Petrologic and geochronologic data on the timing and potential magnitude of associated fluxes call this mechanism into question[15]. Additionally, any estimate of CO_2 release during metamorphism is subject to considerable uncertainty because it is not known what fraction of the CO_2 makes it into the atmosphere after passing through kilometres of variably permeable and reactive rock[15].

Recent modelling work has suggested that the acceleration of continental weathering by Cenozoic uplift was compensated by a decrease by a factor of ~2 in the weathering of ocean island basalts in response to changing climatic conditions[10]. This is not implausible but requires very large changes, and there is currently no direct evidence to test this prediction.

Another possibility that has not yet been systematically explored is that Cenozoic uplift, in addition to increasing rates of CO_2 drawdown by silicate weathering, increased rates of sulphide oxidation coupled to carbonate dissolution, which provided a transient source of CO_2. The oxidation of sulphides (for example, pyrite) produces sulphuric acid following the equation

$$4FeS_2 + 15O_2 + 14H_2O \rightarrow 4Fe(OH)_3 + 8H_2SO_4 \quad (1)$$

The dissolution of carbonate minerals with this sulphuric acid results in the transfer of sedimentary C into the ocean–atmosphere system as CO_2, which can be represented by the equations

$$CaCO_3 + H_2SO_4 \rightarrow CO_2(g) + H_2O + Ca^{2+} + SO_4^{2-} \quad (2)$$

and

$$2CaCO_3 + H_2SO_4 \rightarrow 2Ca^{2+} + SO_4^{2-} + 2HCO_3^{-} \quad (3)$$

For equation (2), the release of CO_2 occurs immediately. For equation (3), CO_2 is still released but on the timescale of carbonate precipitation in the ocean[16] (10^6 yr), represented by the equation

$$Ca^{2+} + 2HCO_3^{-} \rightarrow CaCO_3 + CO_2(g) + H_2O \quad (4)$$

A strong correlation between erosion and sulphide oxidation rates[17] suggests that orogenic activity, in addition to increasing CO_2 drawdown through enhanced silicate weathering, will also increase CO_2 release from coupled sulphide oxidation and carbonate dissolution.

Estimates of the modern sulphide oxidation fluxes range from 0.64 to 1.26 Tmol yr^{-1} and are based on either the chemistry of large rivers[16] or the modelled composition of weathering sources[18]. Direct measurements of sulphide-derived S fluxes have been limited, because isotopic measurements are generally required to distinguish reliably between sulphate and sulphide-derived S. Data from the localities where the relative proportions of sulphide and sulphate-derived S have been constrained indicate a flux of ~0.2 Tmol yr^{-1} of sulphide S, contributed by ~2% of global land area[17,19–22]. Directly extrapolating this value globally would yield a flux of 10 Tmol yr^{-1} sulphide S. Although the areas with well-constrained data may represent environments with disproportionately high sulphide oxidations rates, it is nonetheless likely that the actual global sulphide oxidation rate has been previously underestimated, or at least lies in the upper range of previous estimates.

In the localities where both silicate weathering and sulphide oxidation rates have been measured, it is possible to compare the rates of CO_2 uptake and release (Fig. 1 and Methods). For the Mackenzie River in Canada[17,23] and the Liwu River in Taiwan[20,21], the calculated release rates of CO_2 associated with sulphide oxidation are greater than the uptake rates of CO_2 by silicate weathering, suggesting that these environments release CO_2 into the atmosphere. For the Ganges–Brahmaputra

[1]Department of Earth Sciences, University of Southern California, Los Angeles, California 90089, USA. [2]MOE Key Laboratory of Surficial Geochemistry, Department of Earth Sciences, Nanjing University, 163 Xianlindadao, Nanjing 210046, China.

river system in India[19], the rates of CO_2 consumption and release are comparable, which suggests that all of the CO_2 consumption by silicate weathering is, to first order, balanced by coupled sulphide oxidation and carbonate dissolution.

Ultimately, the CO_2 released by coupled sulphide oxidation and carbonate dissolution is balanced by alkalinity production during sulphate reduction and sulphide precipitation in marine sediments, which can be represented by the generalized equation

$$\begin{aligned}&2FeOOH + 2H^{+} + Ca^{2+} + 2SO_4^{2-} + 4CH_2O \\ &\rightarrow FeS_2 + Fe^{2+} + 4H_2O + Ca^{2+} + 4HCO_3^{-}\end{aligned} \quad (5)$$

The long residence time of sulphate (>10 Myr) relative to Ca^{2+} (1 Myr) in the ocean allows for the transient release of sedimentary C into the atmosphere as a result of coupled sulphide oxidation and carbonate dissolution[16,17].

The controls on the duration of transient CO_2 release can be explored using a conceptual box model of the ocean (Fig. 2 and Methods). In this model, the input of carbon from carbonate weathering by sulphuric acid (equations (1)–(3)) provides an excess sedimentary contribution to the combined ocean–atmosphere pool, represented in our model by dissolved inorganic carbon, or DIC equations (1)–(3). The excess sedimentary DIC is ultimately balanced by DIC derived from the oxidation of organic carbon during sulphate reduction (equation (5)), such that the ratio of sedimentary DIC to total DIC derived from carbonate weathering is equal to 0.5 in steady state. For an instantaneous doubling of weathering-derived input fluxes, an increase in the ratio of sedimentary DIC to total DIC is observed for a period of ~70 Myr, reflecting a transient release of CO_2 (Fig. 2). For a doubling of input fluxes that occurs over a period of 40 Myr, the transient release of CO_2 lasts for ~95 Myr (Fig. 2). These calculated timescales are comparable to the duration of enhanced silicate weathering implied by isotopic records, suggesting that enhanced sulphide oxidation is potentially important in countering the excess CO_2 drawdown caused by tectonic uplift during the Cenozoic.

Sulphide minerals are concentrated in organic-rich sedimentary rocks along with chalcophile elements such as Re and Os (ref. 24). The decay of ^{187}Re to ^{187}Os leads to elevated $^{187}Os/^{188}Os$ ratios in sulphide- and organic-rich sedimentary rocks, making the marine record of $^{187}Os/^{188}Os$ sensitive to the weathering of these lithologies over geologic time[24]. Using the modelling framework of ref. 10, which reconstructs changes

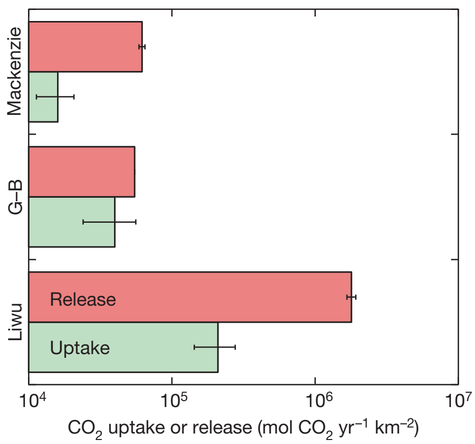


Figure 1 | CO_2 uptake (green bars) and release (red bars) rates calculated for the Mackenzie River[17,22], the Liwu River[20,21] and the Ganges–Brahmaputra (G–B)[19]. The compiled data are presented in Extended Data Table 1, and a description of the calculation method is presented in Methods. Error bars reflect the propagation of the analytical uncertainty of each value that was reported by the original authors (1 s.d.). For the Ganges–Brahmaputra, the uncertainty in the proportion of the sulphate flux derived from sulphide oxidation was not reported and, therefore, no error bar is included.

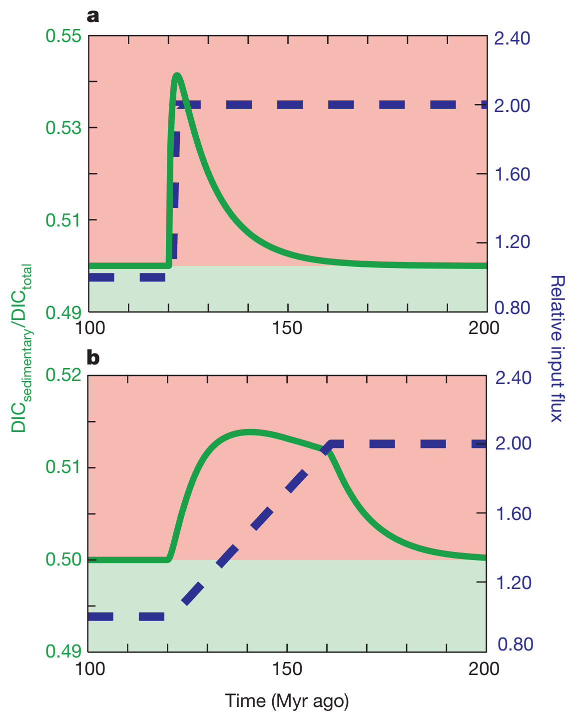


Figure 2 | Conceptual model of the timescale of CO_2 release associated with the contrasting timescales of sulphide and carbonate burial. CO_2 release (red portion of the graph) occurs when the ratio of DIC derived from sedimentary carbonates ($DIC_{sedimentary}$) to total DIC (DIC_{total}) is above 0.5, which is the value associated with the dissolution of carbonates by carbonic acid and is p_{CO_2} neutral (that is, the precipitation of carbonates from a fluid with this composition would have no effect on long-term p_{CO_2}). **a**, The timescale of CO_2 release (solid green line) associated with an instantaneous doubling of the input fluxes (dashed blue line) from carbonate weathering by carbonic acid and sulphuric acid. **b**, The timescale of CO_2 release (solid green line) associated with a doubling of the input fluxes (dashed blue line) from carbonate weathering by carbonic acid and sulphuric acid that occurs progressively over 40 Myr.

in sedimentary rock weathering based in part on an $^{187}Os/^{188}Os$ mass balance, we calculate the CO_2 release as a result of coupled sulphide oxidation and carbonate dissolution during the Cenozoic (Methods). The effects of pyrite burial are not included in the model because there are few constraints on how burial rates may have varied over the Cenozoic.

Using the model output of CO_2 uptake from continental silicate weathering and release related to sulphide oxidation, we calculate the change in the net CO_2 drawdown from mountain uplift throughout the Cenozoic (Fig. 3). The reconstructed net CO_2 drawdown from mountain uplift is relatively constant from 50–25 Myr ago and consistent with proxy-based reconstructions of the atmospheric CO_2 partial pressure[11,12], p_{CO_2} (Fig. 3). From 25–15 Myr ago, there is a period of marked increase in the net CO_2 consumption from mountain uplift that corresponds with a decline in p_{CO_2} (Fig. 3; refs 11, 12). This increase results from a mismatch in the shape of the $^{87}Sr/^{86}Sr$ and $^{187}Os/^{188}Os$ curves such that silicate weathering continues to increase while the weathering of sulphide/organic rich sedimentary rocks remains constant (Fig. 3). After this perturbation, the net CO_2 drawdown from mountain uplift and the p_{CO_2} of the atmosphere both stabilize at respectively higher and lower values relative to 50–25 Myr ago (Fig. 3; refs 11, 12). Lower p_{CO_2} conditions from 15–0 Myr ago would decrease weathering fluxes due to the climate-weathering feedback and act to balance the larger net CO_2 drawdown from mountain uplift over this time (see, for example, ref. 10), which is consistent with $^{10}Be/^{9}Be$ evidence for constant global weathering fluxes over the past 12 Myr (ref. 25). The general consistency between the proxy-based p_{CO_2} record and the model calculations support the hypothesis that sulphide oxidation had an important role in in the Cenozoic C cycle and may help, at least in part, to reconcile radiogenic isotope records

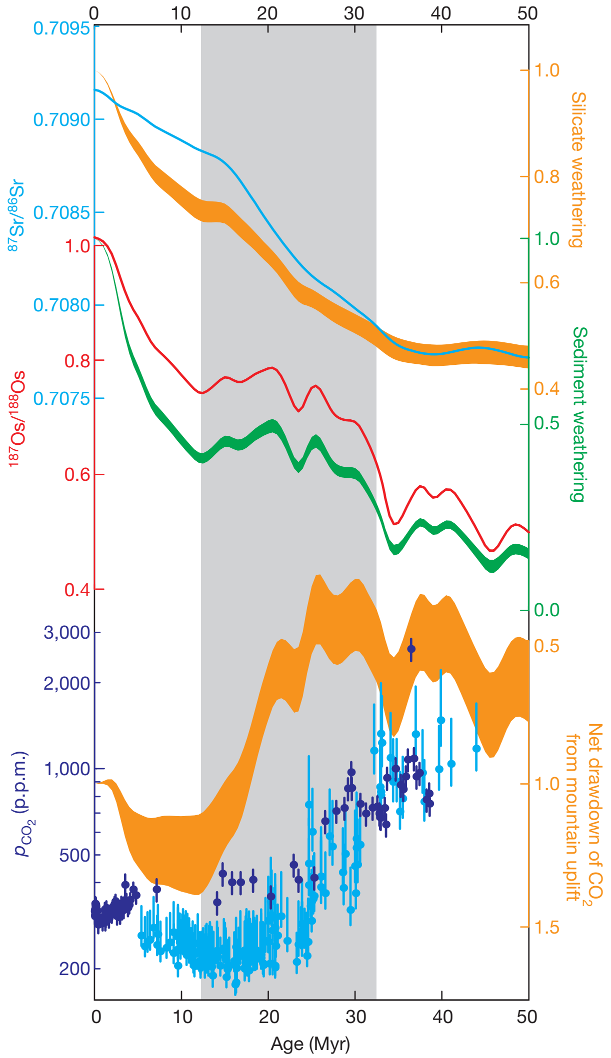


Figure 3 | Comparison of the isotope mass balance model calculations with proxy records for changes in the Cenozoic C cycle. Values of silicate weathering, sedimentary rock weathering, and the net drawdown CO_2 by uplift are all relative to modern fluxes with the width of each of the curves reflecting the propagated uncertainty (1 s.d.) associated with the model input parameters. The shaded period highlights a major increase in the net weathering CO_2 sink that is associated with a drop in atmospheric p_{CO_2} (refs 11, 12). Errors in p_{CO_2} reflect the range of p_{CO_2} values for different proxy calibrations (ref. 11 (light blue) and ref. 12 (dark blue)).

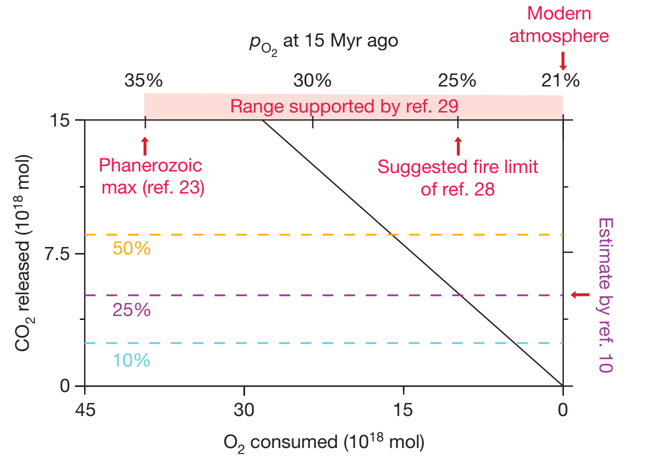


Figure 4 | Trade-off between CO_2 release and O_2 consumption as a result of sulphide oxidation. The dashed horizontal lines represent the CO_2 release required to fully compensate for a 10% (blue), 25% (purple) or 50% (orange) increase in continental silicate weathering that occurred linearly over the past 15 Myr, assuming a modern flux of 3.5×10^{18} mol Myr^{-1} (Methods). The 25% increase corresponds to the increase predicted by the mass balance model of ref. 10 and implies p_{O_2} at 15 Myr ago of ~25%. The estimate for the maximum possible atmosphere O_2 content of 25% (ref. 28) is at odds with modelling[23] and experimental[29] evidence for higher maximum O_2 contents. The solid line shows the ratio of CO_2 release to O_2 consumption for sulphide oxidation.

with atmospheric p_{CO_2} proxies and the requirement of mass balance in the long-term C cycle.

Enhanced sulphide oxidation in response to Cenozoic uplift should also affect the long-term S cycle. Fluid inclusion analyses[26] and the isotopic composition of authigenic barites[27] suggest that marine sulphate concentrations increased during the Cenozoic without a corresponding change in the sulphur isotope ratio. This complicates straightforward interpretation of the Cenozoic S cycle because processes that could modulate marine sulphate concentrations should also affect sulphur isotope ratios. By using an inverse model of the S cycle (Methods), we find that our reconstruction of sedimentary rock weathering can be reconciled with the marine $\delta^{34}S_{sulphate}$ ($\delta^{34}S = ((^{34}S/^{32}S)_{sample}/(^{34}S/^{32}S)_{VCDT} - 1) \times 1{,}000$; VCDT, Vienna-Canyon Diablo Troilite) record for a reasonable range of parameters values (Extended Data Fig. 1), and that the corresponding range of CO_2 release is of similar magnitude to what would be required to balance proposed increases in silicate weathering (Extended Data Fig. 4). Thus, our hypothesis is consistent with the marine $\delta^{34}S_{sulphate}$ record and can simultaneously explain the observed increase in sulphate concentrations through the Cenozoic.

The O_2 content of the atmosphere depends on the balance between the burial and oxidation of reduced carbon and sulphur compounds[23]. An increase in sulphide oxidation will consume O_2, with a molar ratio of O_2 consumption to CO_2 production of 15/8 (equations (1)–(3) and Fig. 4). The hypothesis that CO_2 release as a result of sulphide oxidation has been important in balancing enhanced CO_2 consumption since the Miocene epoch (as suggested by Fig. 3) would require O_2 consumption during this period. Completely compensating for a 25% increase in silicate-weathering CO_2 drawdown over the past 15 Myr (the increase implied in the model of ref. 10 and shown in Fig. 3) would mean an early- to mid-Miocene O_2 partial pressure of $p_{O_2} \approx 25\%$, compared with the present-day value, of 21% (Fig. 4). There is little independent evidence for how much atmospheric p_{O_2} varied during the Cenozoic era, but changes of the order of ~4% are within geologically reasonable ranges[23,28,29] and are consistent with estimates derived from the S cycle modelling (Methods and Extended Data Fig. 4). If O_2 consumption occurred over the whole Cenozoic, as some of our S isotope models suggest (Methods and Extended Data Fig. 1), the additional O_2 demand could require Palaeocene p_{O_2} values of the order of 23–33% (Methods). Importantly, these calculations do not fully capture actual expected changes in p_{O_2}, because the oxygen budget also depends on fluxes in the organic C cycle[23] and the average oxidation state of this carbon, which is variable[30] and poorly constrained. The key point is that the trade-off between O_2 and CO_2 places a practical limit on geologically reasonable CO_2 production by sulphide oxidation, but the relative magnitudes of the reservoirs and fluxes mean that sulphide oxidation can have an important role in balancing silicate-weathering-driven CO_2 consumption if accompanied by modest variation in atmospheric p_{O_2}.

It is unlikely that the evolution of the C cycle during the Cenozoic depended alone on changes in silicate weathering balanced by sulphide–carbonate weathering; additional CO_2 release from metamorphic decarbonation[9,15], decreasing rates of ocean-island-basalt weathering associated with global climate-weathering feedbacks[10], and changes in the organic C cycle may also have been important. Although many questions remain, including about the coincident variation in p_{O_2}, sulphide oxidation coupled to carbonate dissolution does provide an attractive and thus far under-recognized source of CO_2 because it is driven by the same forcing mechanism as enhanced CO_2 consumption by silicate weathering[5,17], operates on a sufficiently long timescale (Fig. 2), and is consistent with proxy

records of weathering fluxes[2–4] and atmospheric p_{CO_2} variations during the Cenozoic (Fig. 3; refs 11, 12).

METHODS SUMMARY

All data for calculating the river fluxes shown in Fig. 1 have been compiled from published literature. The data were corrected for atmospheric inputs and partitioned between lithologic and acid sources using standard methods based on endmember elemental and isotopic ratios.

The conceptual box model presented in Fig. 2 was constructed using a set of ordinary differential equations describing the concentrations of DIC derived from sedimentary rocks, DIC derived from atmospheric CO_2, and sulphate in the ocean considering only the fluxes related to the weathering or precipitation of carbonates and sulphides.

The model[10] used to produce the results shown in Fig. 3 is based on mass balance equations that relate fluxes in the long-term C cycle to isotopic records ($\delta^{13}C$, $^{87}Sr/^{86}Sr$, and $^{188}Os/^{187}Os$), dissolved Mg concentrations[26] and sea-floor spreading rates[6]. Changes in the fluxes relative to modern values are calculated by solving the equations simultaneously using smoothed fits of the proxy data.

For the inverse model of the S cycle, we adopted the approach of ref. 31. Model inputs include the results of the model of ref. 10 (Fig. 3) for the time evolution of the riverine sulphate flux, the marine $\delta^{34}S_{sulphate}$ record of ref. 27 and marine sulphate concentration data from ref. 26. The equation was arranged to solve for a single S cycle parameter while holding the other parameters at constant values reflecting their predicted ranges.

The ratio of O_2 consumption to CO_2 production implied by our model was calculated with equations (1)–(3) and used to assess the Cenozoic O_2 budget for scenarios of enhanced silicate weathering over the past 15 Myr that bracket the range of predicted values[9,10]. Values of O_2 consumption were converted to the p_{O_2} required 15 Myr ago by modelling the atmosphere as an ideal gas (that is, equating the mole fraction with the volume fraction).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 February 2013; accepted 9 January 2014.

1. Walker, J., Hays, P. & Kasting, J. A negative feedback mechanism for the long-term stabilization of the Earth's surface temperature. *J. Geophys. Res.* **86,** 9776–9782 (1981).
2. McArthur, J. M., Howarth, R. J. & Bailey, T. R. Strontium isotope stratigraphy: LOWESS version 3: best fit to the marine Sr-isotope curve for 0–509 Ma and accompanying look-up table for deriving numerical age. *J. Geol.* **109,** 155–170 (2001).
3. Klemm, V., Levasseur, S., Frank, M., Hein, J. R. & Halliday, A. N. Osmium isotope stratigraphy of a marine ferromanganese crust. *Earth Planet. Sci. Lett.* **238,** 42–48 (2005).
4. Misra, S. & Froelich, P. N. Lithium isotope history of Cenozoic seawater: changes in silicate weathering and reverse weathering. *Science* **335,** 818–823 (2012).
5. Raymo, M., Ruddiman, W. & Froelich, P. Influence of late Cenozoic mountain building on ocean geochemical cycles. *Geology* **16,** 649–653 (1988).
6. Müller, R. D., Sdrolias, M., Gaina, C., Steinberger, B. & Heine, C. Long-term sea-level fluctuations driven by ocean basin dynamics. *Science* **319,** 1357–1362 (2008).
7. Berner, R. A. & Caldeira, K. The need for mass balance and feedback in the geochemical carbon cycle. *Geology* **25,** 955 (1997).
8. Raymo, M. & Ruddiman, W. F. Tectonic forcing of late Cenozoic climate. *Nature* **359,** 117–122 (1992).
9. Bickle, M. Metamorphic decarbonation, silicate weathering and the long-term carbon cycle. *Terra Nova* **8,** 270–276 (1996).
10. Li, G. & Elderfield, H. Evolution of carbon cycle over the past 100 million years. *Geochim. Cosmochim. Acta* **103,** 11–25 (2013).
11. Pagani, M., Zachos, J. C., Freeman, K. H., Tipple, B. & Bohaty, S. Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. *Science* **309,** 600–603 (2005).
12. Zhang, Y. G., Pagani, M., Liu, Z., Bohaty, S. M. & DeConto, R. A 40-million-year history of atmospheric CO_2. *Phil. Trans. Royal. Soc. A* **371,** 1–20 (2013).
13. Galy, V., Beyssac, O., France-Lanord, C. & Eglinton, T. Recycling of graphite during Himalayan erosion: a geological stabilization of carbon in the crust. *Science* **322,** 943–945 (2008).
14. Katz, M. *et al.* Biological overprint of the geological carbon cycle. *Mar. Geol.* **217,** 323–338 (2005).
15. Kerrick, D. M. & Caldeira, K. Metamorphic CO_2 degassing from orogenic belts. *Chem. Geol.* **145,** 213–232 (1998).
16. Berner, K. E. & Berner, R. A. *Global Environment: Water, Air and Geochemical Cycles* 369–382 (Princeton Univ. Press, 2012).
17. Calmels, D., Gaillardet, J., Brenot, A. & France-Lanord, C. Sustained sulfide oxidation by physical erosion processes in the Mackenzie River basin: climatic perspectives. *Geology* **35,** 1003 (2007).
18. Lerman, A., Wu, L. & Mackenzie, F. T. CO_2 and H_2SO_4 consumption in weathering and material transport to the ocean, and their role in the global carbon balance. *Mar. Chem.* **106,** 326–350 (2007).
19. Galy, A. & France-Lanord, C. Weathering processes in the Ganges-Brahmaputra basin and the riverine alkalinity budget. *Chem. Geol.* **159,** 31–60 (1999).
20. Calmels, D. *et al.* Contribution of deep groundwater to the weathering budget in a rapidly eroding mountain belt, Taiwan. *Earth Planet. Sci. Lett.* **303,** 48–58 (2011).
21. Das, A., Chung, C.-H. & You, C.-F. Disproportionately high rates of sulfide oxidation from mountainous river basins of Taiwan orogeny: sulfur isotope evidence. *Geophys. Res. Lett.* **39,** L12404 (2012).
22. Millot, R., Gaillardet, J., Dupre, B. & Allegre, C. Northern latitude chemical weathering rates: clues from the Mackenzie River Basin, Canada. *Geochim. Cosmochim. Acta* **67,** 1305–1329 (2003).
23. Berner, R. A. *et al.* Isotope fractionation and atmospheric oxygen: Implications for Phanerozoic O_2 evolution. *Science* **287,** 1630–1633 (2000).
24. Georg, R. B., West, A. J., Vance, D., Newman, K. & Halliday, A. N. Is the marine osmium isotope record a probe for CO_2 release from sedimentary rocks? *Earth Planet. Sci. Lett.* **367,** 28–38 (2013).
25. Willenbring, J. K. & von Blanckenburg, F. Long-term stability of global erosion rates and weathering during late-Cenozoic cooling. *Nature* **465,** 211–214 (2010).
26. Horita, J., Zimmermann, H. & Holland, H. D. Chemical evolution of seawater during the Phanerozoic: implications from the record of marine evaporites. *Geochim. Cosmochim. Acta* **66,** 3733–3756 (2002).
27. Paytan, A., Kastner, M., Campbell, D. & Thiemens, M. H. Sulfur isotopic composition of Cenozoic seawater sulfate. *Science* **282,** 1459–1462 (1998).
28. Watson, A. J., Lovelock, J. E. & Margulis, L. Methanogenesis, fires, and the regulation of atmospheric oxygen. *Biosystems* **10,** 293–298 (1978).
29. Wildman, R. A. *et al.* Burning of forest materials under late Paleozoic high atmospheric oxygen levels. *Geology* **32,** 457–460 (2004).
30. LaRowe, D. E. & Van Cappellen, P. Degradation of natural organic matter: A thermodynamic analysis. *Geochim. Cosmochim. Acta* **75,** 2030–2042 (2011).
31. Kurtz, A. C., Kump, L. R., Arthur, M. A., Zachos, J. C. & Paytan, A. Early Cenozoic decoupling of the global carbon and sulfur cycles. *Paleoceanography* **18,** 1090 (2003).

Acknowledgements Support for this work comes from a USC College Fellowship and a C-DEBI Graduate Fellowship to M.A.T., NSF funding (NSF-EAR/GLD-1053504 and EAR/GLTG-1227192) to A.J.W., and National Natural Science Foundation of China funding (grant nos 41173105, 41102103 and 41321062) to G.L. This is C-DEBI contribution #197.

Author Contributions M.A.T. and A.J.W. designed the study, the conceptual model experiments and the sulphur isotope modelling. G.L. contributed the isotope mass balance model. M.A.T. and A.J.W. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.T. (marktorr@usc.edu), A.J.W (joshwest@usc.edu), or G.L (ligaojun@nju.edu.cn).

METHODS

Calculation of CO_2 fluxes from modern rivers. All data for calculating the river fluxes shown in Fig. 1 have been compiled from published literature. CO_2 uptake rates are calculated from fluxes of Na, K, Ca and Mg corrected for rainwater input using Cl^- concentrations and rainwater cation/Cl^- ratios. To partition the Ca flux between carbonate and silicate components, an endmember silicate Ca/Na ratio of 0.35 was used for the Liwu[20] and Mackenzie[22] and a value of 0.2 was used for the Ganges–Brahmaputra[19] (G–B). To partition the Mg flux between silicate and carbonate sources, an endmember silicate Mg/Na ratio of 0.24 was used for the Liwu[20] and Mackenzie[22] whereas a Mg/K ratio of 0.5 was used for the G–B[19] following the methods of the original authors. All of the Na and K fluxes after correction for rain contributions are assumed to be derived from silicate sources except in the Mackenzie, where a portion of Na flux is contributed by evaporites and was accounted for using $^{87}Sr/^{86}Sr$ ratios assuming an evaporite endmember ratio of 0.7081 (ref. 22). All of the silicate Ca and Mg fluxes were assumed to contribute to long-term CO_2 drawdown whereas only 30% of the Na flux and 20% of the K flux were considered[32]. For the G–B calculations, some of our values differ slightly from those reported in ref. 19, but this does not affect the overall interpretation. The calculated values are presented in Extended Data Table 1.

To calculate CO_2 release rates, the proportion of the sulphate flux derived from sulphide oxidation was estimated using $\delta^{34}S_{sulphate}$ (Liwu[21]), $\delta^{34}S_{sulphate}$ and $\delta^{18}O_{sulphate}$ (Mackenzie[17]) or $\delta^{13}C_{DIC}$ (G–B[19]). All of the sulphide-derived SO_4^{2-} flux was assumed to react preferentially with carbonates, which provides a maximum estimate of CO_2 release rates. The relative difference between release and uptake rates (that is, the comparison shown in Fig. 1) does not change depending on the partitioning of the sulphide-derived SO_4^{2-} between silicate and carbonate rocks. For the Liwu calculation, the $\delta^{34}S_{sulphate}$ measurements used to calculate the proportion of sulphide-derived SO_4^{2-} are from the adjacent Kaoping River[21] because there are no published values from the Liwu. Annual fluxes from the Liwu are considered because they are better constrained, and the two rivers drain similar lithologies[20,21].

Conceptual box model. The conceptual box model (Fig. 2) was constructed using a set of ordinary differential equations describing the concentrations of DIC and sulphate in the ocean considering only the fluxes related to the weathering or precipitation of carbonates and sulphides (equations (1)–(5)). Initial values of 8×10^{18} mol Myr^{-1}, 5×10^{18} mol Myr^{-1} and 1×10^{18} mol Myr^{-1} are used for the input fluxes of total DIC, sedimentary DIC and sulphide-derived sulphate, respectively. The output fluxes (carbonate precipitation and sulphate reduction or sulphide precipitation) are assumed to vary linearly as functions of concentration with rate constants of 1 Myr^{-1} for DIC and 0.1 Myr^{-1} for sulphate[16]. The model was run for 120 Myr to reach a steady state before the input fluxes were doubled instantaneously (Fig. 2a) or over the course of 40 Myr (Fig. 2b).

Li and Elderfield mass balance model. The results from the adapted Li and Elderfield[10] model shown in Fig. 3 are based on a series of isotope mass balance equations that relate the weathering of continental silicates, the weathering of sedimentary rocks (which includes organic C (C_{org}) and carbonate weathering), the weathering of basaltic rocks, reverse weathering or hydrothermal alteration, C_{org} burial, carbonate burial, and volcanic CO_2 degassing to the marine records of $^{87}Sr/^{86}Sr$ (ref. 2), $^{187}Os/^{188}Os$ (ref. 3), $\delta^{13}C_{org}$ (ref. 33), $\delta^{13}C_{carbonate}$ (ref. 33), dissolved Mg concentrations[26] and sea-floor spreading rates[6]. The mass balance equations are solved simultaneously using smoothed fits of the published marine proxy records to calculate changes in the fluxes of the different C cycle parameters relative to modern values (the modern values used in the calculation are presented in ref. 10).

To simplify the model, various assumptions are made regarding the mass balances of each element or isotope system. For the Os system, the isotopic composition of the ocean is considered to be controlled by the weighted average of inputs from continental silicate rocks, island basalts, sedimentary rocks, hydrothermal activity and cosmic dust as described by the equation

$$\sum_i [K_i \mathrm{Os}_i (R_i^{\mathrm{Os}} - R^{\mathrm{Os}}_{\mathrm{ocean}})] = 0 \qquad (6)$$

where K_i is the intensity of the ith flux relative to the late Pleistocene value, Os_i is the late Pleistocene value of the ith flux, R_i^{Os} is the normalized osmium isotopic composition of the ith flux and $R^{\mathrm{Os}}_{\mathrm{ocean}}$ is the normalized osmium isotopic composition of the ocean[10]. This is a simplification of the Os system because different sedimentary lithologies contribute different amounts of Os with distinct $^{188}Os/^{187}Os$ values. For example, studies of modern river systems[34,35] have suggested that carbonates contribute non-radiogenic Os whereas organic-rich shales contribute more radiogenic Os. This is taken into account in our model by assigning a mean $^{188}Os/^{187}Os$ value for sedimentary rocks that reflects a weighted contribution from different sedimentary lithologies[10,36]. Because this value is held constant over the Cenozoic, we assume that weathering fluxes from different sedimentary lithologies change in a constant proportion. This is a reasonable assumption because different sedimentary lithologies are interbedded on a large scale, and thus exhumed in tandem. If this proportion were to vary, one would predict that Os fluxes from carbonates would increase relative to organic-rich shales because carbonate minerals are more abundant[18]. This would result in an underestimation of changes in sedimentary rock weathering making our calculation of CO_2 release from sulphide oxidation a minimum estimate, thereby strengthening our argument about the potential importance of this mechanism. The Os system is further simplified by assuming a constant cosmic input flux.

Owing to the long residence time of Sr relative to the modelling timescale, the transient-state solution to the mass balance equation is used, and expressed as

$$N \mathrm{d}R^{\mathrm{Sr}}_{\mathrm{ocean}}/\mathrm{d}t = \sum_i [K_i \mathrm{Sr}_i (R_i^{\mathrm{Sr}} - R^{\mathrm{Sr}}_{\mathrm{ocean}})] \qquad (7)$$

where N is the amount of Sr in seawater, $\mathrm{d}R^{\mathrm{Sr}}/\mathrm{d}t$ is the time derivative of normalized Sr isotopic composition of the ocean, Sr_i is the late Pleistocene value of the ith flux, R_i^{Sr} is the normalized Sr isotopic composition of the ith flux and $R^{\mathrm{Sr}}_{\mathrm{ocean}}$ is the normalized Sr isotopic composition of the ocean[10,36]. In this mass balance, contributions from continental silicate rocks, island basalts, sedimentary rocks (that is, carbonates), hydrothermal activity and sediment diagenesis are considered. Like the Os isotope system, the reconstruction of weathering fluxes using the marine $^{87}Sr/^{86}Sr$ record is sensitive to the isotopic composition of source rocks. Although carbonate minerals typically have low $^{87}Sr/^{86}Sr$ values relative to igneous rocks, intense metamorphic activity, such as that experienced in the Himalayan region, can impart unusually radiogenic Sr isotope compositions to carbonate minerals[37]. The potential for radiogenic Sr fluxes from metamorphic carbonates in the Himalayan region to bias our reconstruction of silicate weathering fluxes was previously assessed by Li *et al.*[36] and Li and Elderfield[10]. This sensitivity analysis was conducted by modelling a linear increase in the isotopic composition of the Sr flux from Tibetan rivers (from 0.7077 to 0.7081) over the past 40 Myr and partitioning this flux in variable proportions between 100% carbonate and 100% silicate sources to test the effects of increasing the input of radiogenic Sr from Tibetan rivers over the Cenozoic on the reconstruction of weathering fluxes. Consideration of these endmember scenarios minimally affected the reconstructed silicate weathering fluxes[10,36].

To incorporate CO_2 release due to coupled sulphide oxidation and carbonate dissolution into this model, modern rates of sulphide oxidation were taken to be equal to 50% of the modern natural riverine S flux[16] and kept in a constant proportion to the total weathering flux of sedimentary rocks, which is calculated on the basis of the mass balances of $^{87}Sr/^{86}Sr$ (equation (6)), $^{187}Os/^{188}Os$ (equation (7)), $\delta^{13}C$ and alkalinity, but is most sensitive to the $^{187}Os/^{188}Os$ mass balance. The net weathering CO_2 drawdown associated with uplift (Fig. 3) is calculated by taking the difference between the CO_2 consumption by silicate weathering and CO_2 release from coupled carbonate dissolution and sulphide oxidation. This approach assumes that a majority of sulphides are hosted in sedimentary rocks because it neglects contributions from igneous sulphide minerals. If igneous sulphide oxidation is a globally important process, our model provides an upper estimate of the net weathering CO_2 drawdown associated with uplift because it underestimates the CO_2 flux from sulphide oxidation.

Sulphur isotope mass balance model. To determine how the pyrite burial flux, the $\delta^{34}S$ of riverine input, and the isotopic fractionation factor associated with sulphate reduction and pyrite burial ($\Delta_{\text{pyrite-seawater}}$) would have to vary to reconcile the marine $\delta^{34}S$ record with the reconstruction of sedimentary rock weathering from the Li and Elderfield[10] model, the non-steady-state inverse-modelling approach of ref. 31 was used. In this approach, the differential equation describing the isotopic mass balance of S is

$$\mathrm{d}\delta^{34}\mathrm{S}/\mathrm{d}t = (F_{\mathrm{riverine}}(\delta^{34}\mathrm{S}_{\mathrm{riverine}} - \delta^{34}\mathrm{S}_{\mathrm{ocean}}) - F_{\mathrm{pyrite\,burial}}\Delta_{\text{pyrite-seawater}})/M_{\mathrm{s}} \qquad (8)$$

where $\mathrm{d}\delta^{34}\mathrm{S}/\mathrm{d}t$ is the time rate of change of the isotopic composition of marine sulphate, $F_{riverine}$ is the input flux of S from rivers (10^{18} mol Myr^{-1}), $\delta^{34}S_{riverine}$ is the isotopic composition of the riverine input, $\delta^{34}S_{ocean}$ is the isotopic composition of marine sulphate, $F_{\mathrm{pyrite\,burial}}$ is the burial flux of S as pyrite (10^{18} mol Myr^{-1}), $\Delta_{\text{pyrite-seawater}}$ is the average fractionation factor associated with sulphate reduction and pyrite burial, and M_s is the mass of marine sulphate (10^{18} mol). We rearrange equation (8) to calculate the time evolution of the parameter of interest. To constrain the time variation of $F_{riverine}$, we use our reconstruction of sedimentary rock weathering from the Li and Elderfield[10] mass balance model (Fig. 3), assuming that the total S yield (that is, evaporite plus sulphide) from sedimentary rock weathering has remained constant and that the modern-day riverine sulphate flux is equal to 3×10^{18} mol S Myr^{-1} (ref. 16). By modelling the riverine input flux in this manner, we are assuming that both evaporite and sulphide oxidation fluxes increased in concert over the Cenozoic in response to tectonic uplift, which is geologically

reasonable because different sedimentary lithologies are interbedded on a large scale. As a result, the isotopic composition of the riverine input does not have to vary over the Cenozoic in our model. Nevertheless, the potential for variation in the isotopic composition of the riverine input to influence the Cenozoic S cycle was explored using the inverse-model calculations (see below).

To constrain the time variation in the mass of marine sulphate, we use an exponential fit to the fluid inclusion data from ref. 26, following ref. 31. To constrain the time-dependent term $d\delta^{34}S/dt$, we calculate the first derivative of a smoothing spline fit of the $\delta^{34}S_{barite}$ record[27] after a 20-point robust loess smoothing of the raw data. This fit to the $\delta^{34}S_{barite}$ record is also used for the value of $\delta^{34}S_{ocean}$ in the model.

The time variations of the fractionation factor associated with sulphate reduction and pyrite burial ($\Delta_{pyrite\text{-}seawater}$), the S isotopic composition of riverine input ($\delta^{34}S_{riverine}$) and the pyrite burial flux ($F_{pyrite\ burial}$) are all poorly constrained. As a result, it is necessary to perform separate model experiments where two of the parameters values are fixed at constant values and the third is calculated. In this study, three types of model experiments were conducted, (1) variable pyrite burial fluxes with constant $\delta^{34}S_{riverine}$ and $\Delta_{pyrite\text{-}seawater}$, (2) variable $\delta^{34}S_{riverine}$ with constant $F_{pyrite\ burial}$ and $\Delta_{pyrite\text{-}seawater}$, and (3) variable $\Delta_{pyrite\text{-}seawater}$ with constant $F_{pyrite\ burial}$ and $\delta^{34}S_{riverine}$. To calculate the implied changes in the pyrite burial flux (model type (1)), $\Delta_{pyrite\text{-}seawater}$ was varied between -50 and -30‰ following Ref. 38 and the $\delta^{34}S_{riverine}$ input was varied between 0 and +10‰ following Refs. 27,31. It is important to note that this range of $\delta^{34}S$ values for the riverine input does not necessarily reflect the balance between evaporite and sulphide sources, because each source displays large isotopic variability[38] and riverine sulphate may be fractionated by biological processes before being discharged into the ocean[39]. To calculate the implied changes in $\Delta_{pyrite\text{-}seawater}$ (model type (2)), the $\delta^{34}S$ of riverine input was varied between 0 and +10‰ and the pyrite burial flux was kept at a constant value of 0.22×10^{18}, 0.67×10^{18} or 1.0×10^{18} mol S Myr^{-1}, following refs 27, 31. To calculate the implied changes in the $\delta^{34}S$ of riverine input (model type (3)), $\Delta_{pyrite\text{-}seawater}$ was varied between −50% and −30‰ and the pyrite burial flux was kept at a constant value of 0.22×10^{18}, 0.67×10^{18} or 1.0×10^{18} mol S Myr^{-1}. The constant pyrite burial fluxes greater than 0.22×10^{18} mol S Myr^{-1} used in the model calculations of $\Delta_{pyrite\text{-}seawater}$ and $\delta^{34}S_{riverine}$ can be greater than the riverine input flux reconstructed from the Li and Elderfield[10] isotope mass balance for a few time periods between 30 and 50 Myr ago. For this reason, the model experiments for variable $\delta^{34}S$ of riverine input and $\Delta_{pyrite\text{-}seawater}$ were evaluated only for the period 0–30 Myr ago.

Solving the sulphur mass balance model for variable pyrite burial flux. For the model experiments where the mass balance equation was arranged to solve for the pyrite burial flux with constant values for the $\delta^{34}S$ of riverine input and $\Delta_{pyrite\text{-}seawater}$, the calculated pyrite burial fluxes were directly proportional to the riverine input flux. The constants of proportionality varied depending on the $\delta^{34}S$ of riverine input and $\Delta_{pyrite\text{-}seawater}$. Lower values of the $\delta^{34}S$ of riverine input and higher values of $\Delta_{pyrite\text{-}seawater}$ produced higher pyrite burial fluxes for the same riverine input flux (Extended Data Fig. 1). Multiple model experiments were consistent with the long-term average burial flux estimated in ref. 40 from the pyrite content of marine sediments (Extended Data Fig. 1). For the period around 45–50 Myr ago, where there is a major positive excursion in the marine $\delta^{34}S$ record[27], some of the model experiments produced pyrite burial fluxes greater than the total riverine input flux (Extended Data Fig. 1). After this brief period, all of the model experiments produced burial fluxes that were lower than the total riverine input flux (Extended Data Fig. 1). For model experiments where the $\delta^{34}S$ of riverine input was equal to 10‰, calculated pyrite burial fluxes were less than the estimated sulphide oxidation fluxes used in the Li and Elderfield[10] mass balance model (that is, 50% of the total S flux) regardless of the value of $\Delta_{pyrite\text{-}seawater}$ (Extended Data Fig. 1). For model experiments where the $\delta^{34}S$ of riverine input was equal to 5‰, calculated pyrite burial fluxes were less than the estimated sulphide oxidation fluxes for $\Delta_{pyrite\text{-}seawater}$ values <−30‰ (Extended Data Fig. 1).

For model experiments where the $\delta^{34}S$ of riverine input was equal to 0‰, calculated pyrite burial fluxes were less than the estimated sulphide oxidation fluxes for $\Delta_{pyrite\text{-}seawater}$ values <−40‰ (Extended Data Fig. 1). For reference, the value of $\Delta_{pyrite\text{-}seawater}$ estimated from the average isotopic offset between coeval pyrite and marine sulphate for the past 200 Myr is -43 ± 2‰ (ref. 38).

Solving the sulphur mass balance model for variable $\Delta_{pyrite\text{-}seawater}$. For the model experiments where the mass balance equation was arranged to solve for $\Delta_{pyrite\text{-}seawater}$ with constant values for the pyrite burial flux and the $\delta^{34}S$ of riverine input, higher $\delta^{34}S$ values of riverine input and lower burial fluxes resulted in lower $\Delta_{pyrite\text{-}seawater}$ values for the same input flux (Extended Data Fig. 2). High burial fluxes or high values of $\delta^{34}S_{riverine}$, or both, were required to maintain $\Delta_{pyrite\text{-}seawater}$ values above −60‰, which has been suggested to be the maximum extent of isotopic fractionation during the Phanerozoic eon[38,41]. All model experiments produced values that were distinct from estimates based on the measured average isotopic offset between coeval pyrite and marine sulphate for at least a portion of the period from 0 to 30 Myr ago[38]. From 0 to 12 Myr ago, each model experiment showed decreasing $\Delta_{pyrite\text{-}seawater}$ values (Extended Data Fig. 2).

Solving the sulphur mass balance model for variable $\delta^{34}S_{riverine}$. For the model experiments where the mass balance equation was arranged to solve for the $\delta^{34}S$ of riverine input with constant values for the pyrite burial flux and $\Delta_{pyrite\text{-}seawater}$, larger values of $\Delta_{pyrite\text{-}seawater}$ and higher burial fluxes resulted in lower $\delta^{34}S$ values of riverine input for the same input flux (Extended Data Fig. 3). From 0 to 12 Myr ago, model experiments where the burial flux was $\geq 0.67 \times 10^{18}$ mol S Myr^{-1} showed increasing values of the $\delta^{34}S$ of riverine input (Extended Data Fig. 3). These experiments also produced calculated $\delta^{34}S$ values of riverine input at 0 Myr ago near estimates for the modern value[27,31].

Role of increased input fluxes of sulphur to the oceans. Increasing Cenozoic marine sulphate concentrations[26] require variation in the input or output flux of sulphate, or both. Because the variation of a single parameter within the S cycle would result in a change in the isotopic composition of marine sulphate, covariation of different parameters during the Cenozoic is necessary to reconcile the relative constancy of the marine $\delta^{34}S$ record[27] with increasing marine sulphate concentrations[26]. The results presented here suggest that for a reasonable range of parameter values, increasing marine sulphate concentrations during the Cenozoic could result from increasing input fluxes from weathering if pyrite burial rates, $\Delta_{pyrite\text{-}seawater}$, or the $\delta^{34}S$ of riverine input, or all three, covary with the input flux (Extended Data Figs 1–3). Despite this agreement for some of the model experiments, a mechanism for the covariation with the riverine input flux is required.

The factors that control the fractionation factor associated with sulphate reduction and pyrite burial are incompletely understood, but $\Delta_{pyrite\text{-}seawater}$ is thought to be positively correlated with sulphate concentrations[42] (that is, higher $[SO_4^{2-}]$ leads to less isotopic fractionation). This behaviour is inconsistent with our model because a negative correlation between sulphate concentrations and $\Delta_{pyrite\text{-}seawater}$ is necessary for the $\delta^{34}S$ record to be reconciled with increasing input fluxes by changing only $\Delta_{pyrite\text{-}seawater}$ (Extended Data Fig. 2). Additionally, the model calculations of $\Delta_{pyrite\text{-}seawater}$ are not in good agreement with independent estimates from the sedimentary record[38]. As a result, it is unlikely that variation in $\Delta_{pyrite\text{-}seawater}$ alone can explain variations in the Cenozoic S cycle.

Model experiments where the burial flux of S was $\geq 0.67 \times 10^{18}$ mol S Myr^{-1} produced $\delta^{34}S_{riverine}$ values in agreement with independent estimates[27,31]. In these experiments, the relative constancy of the $\delta^{34}S$ of marine sulphate was partly maintained in spite of increasing input fluxes through an increase in the $\delta^{34}S$ value of riverine input from 0 to 12 Myr ago (Extended Data Fig. 3). This increase could arise from a change in the source of sulphate in response to tectonic uplift. An increase in the portion of sulphate derived from evaporite weathering from <50% during a majority of the Cenozoic to the modern value of ~50% starting at ~12 Myr ago would sufficiently increase the $\delta^{34}S$ value of riverine input. Such an increase in evaporite weathering would mean that we have underestimated the effects of sulphide oxidation in our isotope mass balance model (Fig. 3), because we assume that the portion of riverine sulphate derived from sulphide oxidation remained at the modern value for the entirety of the Cenozoic when it may have been greater during the early Cenozoic.

It is also possible to increase the $\delta^{34}S$ value of riverine input without changing the proportion of sulphide-derived sulphate, because sulphide minerals themselves display a wide range of isotopic values depending on the depositional environment and the geologic age[38]. Similarly, a greater flux of sulphate from environments where the microbial processing of riverine sulphate in groundwaters is important (for example the Himalayan system[39]) could produce an increase in the $\delta^{34}S$ value of riverine input without changing the proportion derived from sulphide oxidation.

Increasing pyrite burial fluxes could also reconcile the Cenozoic marine $\delta^{34}S$ curve with enhanced riverine fluxes and increasing sulphate concentrations (Extended Data Fig. 1). This linkage is consistent with the known coupling between organic C and pyrite burial fluxes[43] and the importance of physical erosion in driving enhanced rates of sulphide oxidation and organic C burial[17,44]. However, a tight coupling between riverine input and pyrite burial fluxes could potentially limit the timescale of transient CO_2 release in response to sulphide oxidation. The sensitivity analysis used in this study suggests that for a reasonable range of $\delta^{34}S$ values of riverine input and $\Delta_{pyrite\text{-}seawater}$ (see above), sulphide oxidation fluxes can remain in excess of pyrite burial fluxes throughout the Cenozoic, which is required for transient CO_2 release (Extended Data Fig. 1). Although the magnitude of CO_2 release from sulphide oxidation will depend on the difference between pyrite oxidation and burial fluxes, the overall increase in CO_2 release during the Cenozoic is preserved. This maintains the general trend of changes in the net weathering CO_2 sink from tectonic uplift throughout the Cenozoic calculated with the Li and Elderfield[10] mass balance model (Fig. 3).

Role of decreased pyrite burial fluxes. Apart from an increase in the riverine input flux, a long-term decrease in pyrite burial fluxes could also explain increasing

Cenozoic marine sulphate concentrations. Unlike increasing riverine fluxes in response to tectonic uplift, it is unclear what processes could lead to a decrease in pyrite burial fluxes and not affect the isotopic composition of marine sulphate. However, it is important to note that decreasing pyrite burial fluxes over the Cenozoic would lead to an imbalance in the S cycle and CO_2 release into the ocean–atmosphere system in the same manner as enhanced continental sulphide oxidation.

Summary of the S cycle modelling. The plausible variation in the $\delta^{34}S$ of riverine input and pyrite burial fluxes determined for a reasonable range of parameter values by our model calculations supports the idea that sulphide oxidation was enhanced during the Cenozoic in response to tectonic uplift. In addition to providing a means of reconciling our hypothesis with the marine $\delta^{34}S$ record, these results also help to explain the observed increase in marine sulphate concentrations during the Cenozoic[26]. Regardless of the specific driver, the modelled scenarios produce conditions under which CO_2 is released into the ocean–atmosphere system in response to sulphide oxidation fluxes in excess of sulphide burial fluxes.

Determining the effect of enhanced sulphide oxidation on Cenozoic p_{O_2}. To assess the trade-offs between O_2 consumption and CO_2 production implied by our model, we considered a molar ratio of O_2 consumption to CO_2 production for pure sulphide oxidation of 15/8, which is based on equations (1)–(3) and is consistent with previous work quantifying the effect of sulphur redox transformations on atmospheric oxygen[23]. We have proposed that sulphide oxidation may have had an important role in balancing enhanced CO_2 consumption by silicate weathering over the last 15 Myr (see, for example, Fig. 3), and so we considered the specific case of the O_2–CO_2 trade-off over this time interval (Fig. 4). We converted the O_2 consumption values to the required p_{O_2} at 15 Myr by using a value of 37.8×10^{18} mol for the present-day mass of O_2 in the ocean–atmosphere system and treating the atmosphere as an ideal gas (that is, equating the mole fraction with the volume fraction). To compare these p_{O_2} consumption values with the CO_2 fluxes, CO_2 drawdown from silicate weathering was calculated for linear increases of 10%, 25% and 50% from 15 Myr ago to modern-day values. The range corresponding to increases of 10–50% brackets the range of previous predictions[9,10] and includes the values implied by Fig. 3. A modern-day silicate weathering flux of 3.5×10^{18} mol Myr was calculated from the flux data of refs 41, 45 and the global lithologic partitioning of ref. 41, assuming that 100% of the riverine silicate Ca^{2+} and Mg^{2+} flux contributed to CO_2 drawdown and that 30% of the Na^+ flux and 20% of the K^+ flux contributed after exchanging for Ca^{2+} in marine sediments (following the approach of ref. 32, and as described above). The implied p_{O_2} at 15 Myr ago, as considered in this analysis, can be readily compared to the fire limit of 25% O_2 suggested by ref. 28, although the experiments of ref. 29 suggest that the fire limit is above 35% O_2, and the modelling results of ref. 23 suggest a Phanerozoic maximum value of $\sim$35% O_2.

To determine the O_2 consumption implied by the S cycle inverse models (Extended Data Fig. 4), the net sulphur flux was calculated by subtracting the sulphide burial flux from the sulphide oxidation flux at million-year intervals. The net sulphide flux was integrated over the period from either 0 to 50 Myr ago or 0 to 15 Myr ago after performing a cubic interpolation using MATLAB. The integrated value was converted from moles S to moles O_2 consumption using a ratio of O_2 consumption to sulphide oxidation of 15/8 (equation (1)). Implied Palaeocene and Miocene p_{O_2} values were determined, as above, on the basis of a present-day O_2 inventory in the atmosphere of 37.8×10^{18} mol, and by treating the atmosphere as an ideal gas (that is, equating mole fraction with the volume fraction).

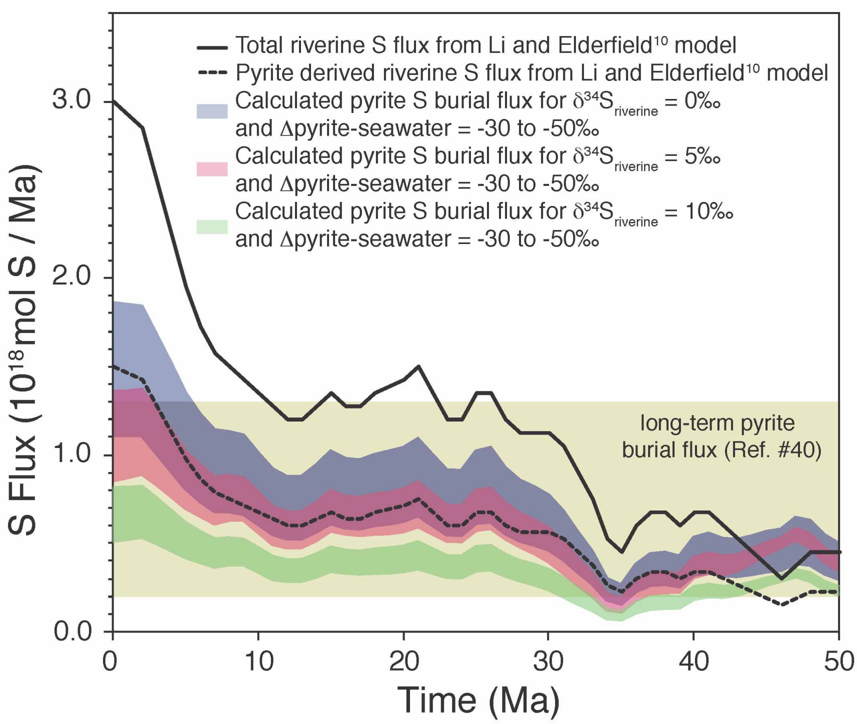
Implications of enhanced sulphide oxidation for Cenozoic p_{O_2}. Integrated over the past 50 Myr, the O_2 consumption implied by the results of the models with variable sulphide burial varies from -26.8×10^{18} to 27.5×10^{18} mol O_2, where negative values reflect net O_2 production (Extended Data Fig. 4a). Net oxygen production is a consequence of calculated pyrite burial fluxes that are higher than the sulphide oxidation fluxes used as model input (see above). Integrated over the past 15 Myr, O_2 consumption varies from -9.9×10^{18} to 14.5×10^{18} mol O_2 (Extended Data Fig. 4b).

For all of the model experiments where pyrite burial fluxes were fixed at constant values, O_2 consumption integrated over the entire period from 0 to 50 Myr ago is 34.1×10^{18}, -7.9×10^{18} or -38.8×10^{18} mol O_2 for respective burial fluxes of 0.22×10^{18}, 0.67×10^{18} and 10^{18} mol S Myr^{-1} (Extended Data Fig. 4a). Integrated over the period from 0 to 15 Myr ago, O_2 consumption is 19.3×10^{18}, 6.7×10^{18} or -2.6×10^{18} mol O_2 for respective burial fluxes of 0.22×10^{18}, 0.67×10^{18} and 10^{18} mol S Myr^{-1} (Extended Data Fig. 4b).

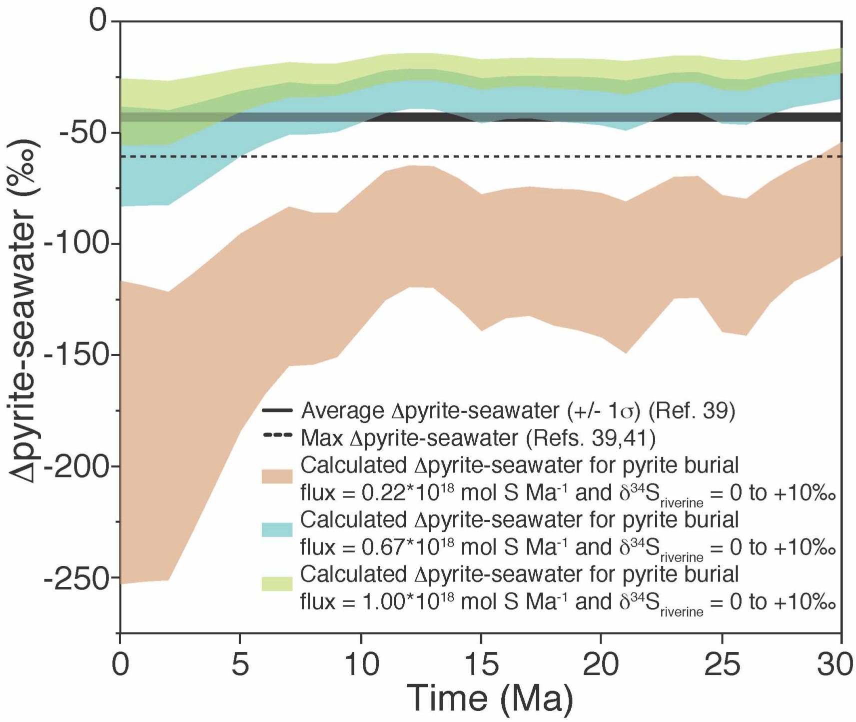
The results of the S cycle model imply changes in atmospheric O_2. Although they notably do not include consideration of the organic C cycle, which may also influence p_{O_2}, the results of the S cycle model imply Palaeocene p_{O_2} values of 23–33% for the combinations of parameter values that produce net O_2 consumption when integrated over the entire period from 0 to 50 Myr ago (Extended Data Fig. 4a). Arguably, the combinations of parameter values that produce net O_2 consumption together with CO_2 production (that is, a $\delta^{34}S$ of riverine input >0‰ and $\Delta_{\text{pyrite-seawater}} \leq -40$) are the most geologically reasonable given that the $\delta^{34}S$ of riverine input is thought to be between +3 and +8‰ (refs 40, 46), and the average $\Delta_{\text{pyrite-seawater}}$ is thought to be -43 ± 2‰ (ref. 38).

Considering only the past 15 Myr, which is when the Li and Elderfield[10] model suggests that sulphide oxidation acted to balance enhanced CO_2 consumption as a result of silicate weathering, the parameter combinations that produce net sulphide oxidation imply Miocene p_{O_2} values of less than 30% and would produce sufficient CO_2 to balance plausible increases in silicate weathering fluxes (Extended Data Fig. 4b). This is consistent with the analysis described in Fig. 4.

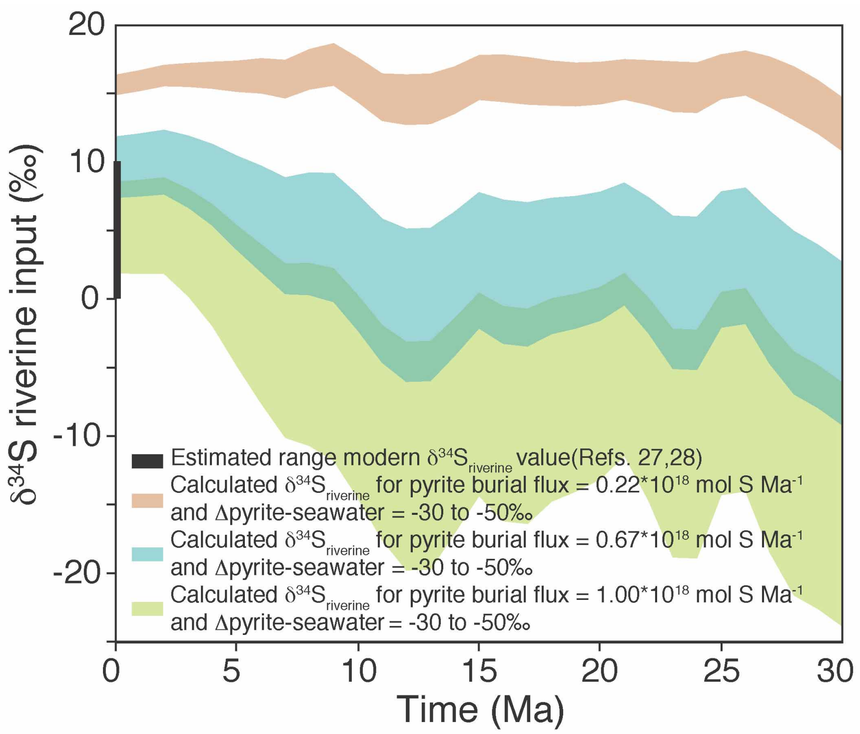
32. France-Lanord, C. & Derry, L. Organic carbon burial forcing of the carbon cycle from Himalayan erosion. *Nature* **390,** 65–67 (1997).
33. Falkowski, P. G. *et al.* The rise of oxygen over the past 205 million years and the evolution of placental mammals. *Science* **309,** 2202–2204 (2005).
34. Huh, Y., Birck, J.-L. & Allègre, C. J. Osmium isotope geochemistry in the Mackenzie River basin. *Earth Planet. Sci. Lett.* **222,** 115–129 (2004).
35. Levasseur, S., Birck, J. & Allegre, C. The osmium riverine flux and the oceanic mass balance of osmium. *Earth Planet. Sci. Lett.* **174,** 7–23 (1999).
36. Li, G., Ji, J., Chen, J. & Kemp, D. B. Evolution of the Cenozoic carbon cycle: the roles of tectonics and CO_2 fertilization. *Glob. Biogeochem. Cycles* **23,** GB1009 (2009).
37. Edmond, J. Himalayan tectonics, weathering processes, and the strontium isotope record in marine limestones. *Science* **258,** 1594–1597 (1992).
38. Wu, N., Farquhar, J., Strauss, H., Kim, S.-T. & Canfield, D. E. Evaluating the S-isotope fractionation associated with Phanerozoic pyrite burial. *Geochim. Cosmochim. Acta* **74,** 2053–2071 (2010).
39. Turchyn, A. V., Tipper, E. T., Galy, A., Lo, J.-K. & Bickle, M. J. Isotope evidence for secondary sulfide precipitation along the Marsyandi River, Nepal, Himalayas. *Earth Planet. Sci. Lett.* **374,** 36–46 (2013).
40. Holser, W. T., Schidlowski, M., Mackenzie, F. T. & Maynard, J. B. in *Chemical Cycles and the Evolution of Earth* (eds Gergor, C. B., Garrels, R. M., Mackenzie, F. T., & Maynard, J. B.) 105–173 (Wiley, 1988).
41. Gaillardet, J., Dupré, B., Louvat, P. & Allegre, C. J. Global silicate weathering and CO2 consumption rates deduced from the chemistry of large rivers. *Chem. Geol.* **159,** 3–30 (1999).
42. Habicht, K. S., Gade, M., Thamdrup, B., Berg, P. & Canfield, D. E. Calibration of sulfate levels in the Archean ocean. *Science* **298,** 2372–2374 (2002).
43. Berner, R. Burial of organic carbon and pyrite sulfur in the modern ocean: its geochemical and environmental significance. *Am. J. Sci.* **282,** 451–473 (1982).
44. Galy, V. *et al.* Efficient organic carbon burial in the Bengal fan sustained by the Himalayan erosional system. *Nature* **450,** 407–410 (2007).
45. Meybeck, M. & Ragu, A. in *Environment Information and Assessment* 1–245 (UN Environment Programme, 1996).
46. Ivanov, M. V. in *The Global Biogeochemical Sulphur Cycle* (eds Ivanov, M. V. & Freney, J. R.) 297–356 (Wiley, 1983).



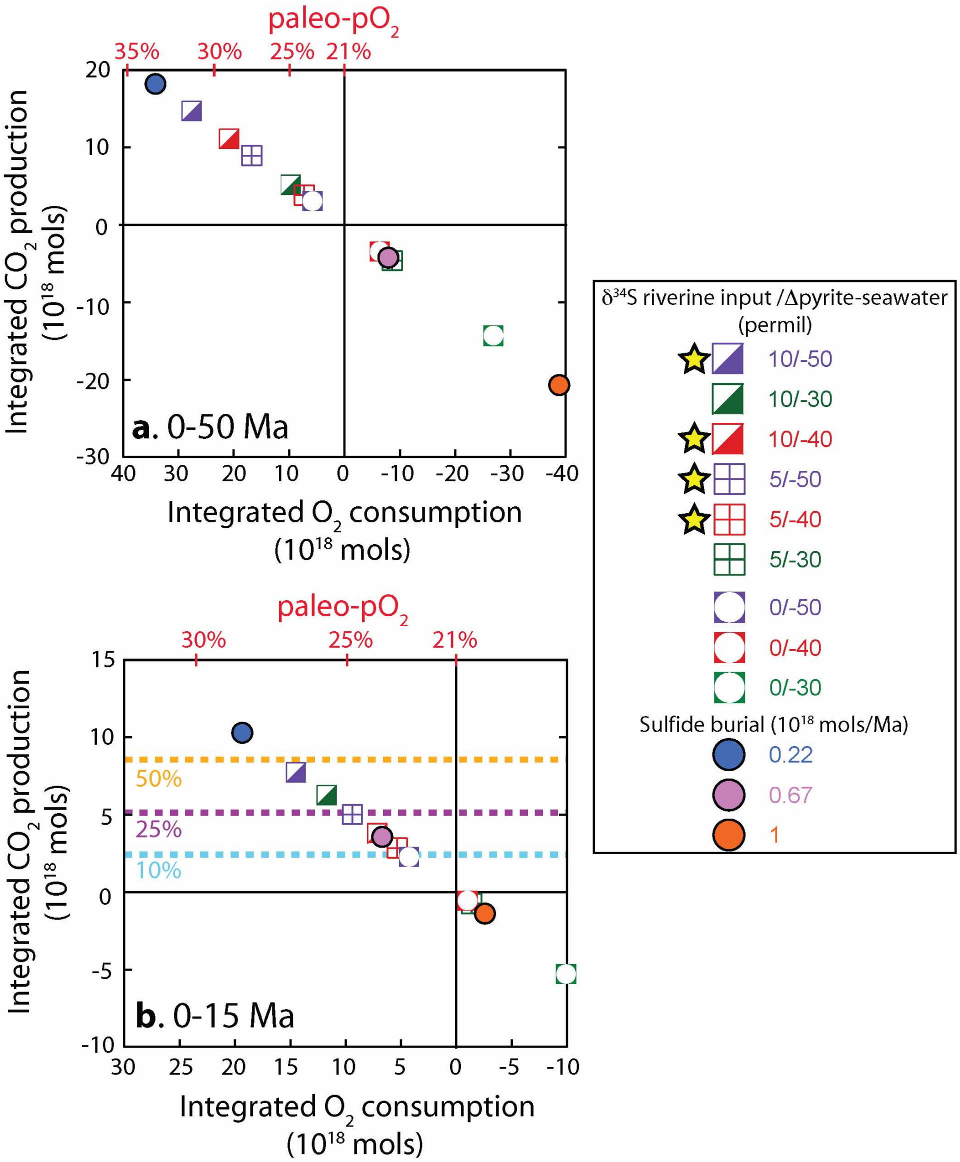
Extended Data Figure 1 | Riverine input fluxes and calculated pyrite burial fluxes from 0 to 50 Myr ago for constant δ^{34}S values of riverine input and $\Delta_{\text{pyrite-seawater}}$. Solid and dashed black lines indicate the total and pyrite-derived input fluxes of sulphate from the Li and Elderfield[10] model, respectively. The green band indicates the range of calculated pyrite burial fluxes for experiments where the δ^{34}S value of riverine S flux was equal to 10‰ and $\Delta_{\text{pyrite-seawater}}$ was varied between −30 and −50‰. The pink band indicates the range of calculated pyrite burial fluxes for experiments where the δ^{34}S value of riverine S flux was equal to 5‰ and $\Delta_{\text{pyrite-seawater}}$ was varied between −30 and −50‰. The blue band indicates the range of calculated pyrite burial fluxes for experiments where the δ^{34}S value of riverine was equal to 0‰ and $\Delta_{\text{pyrite-seawater}}$ was varied between −30 and −50‰. The khaki bar indicates the range of pyrite burial fluxes estimated from the pyrite content of marine sediments by ref. 40.



Extended Data Figure 2 | Calculated $\Delta_{\text{pyrite-seawater}}$ values from 0 to 30 Myr ago for constant δ^{34}S values of riverine input and pyrite burial fluxes. The solid and dashed black lines indicate the average and minimum $\Delta_{\text{pyrite-seawater}}$ values for the Phanerozoic based on the isotopic offset between coeval marine sulphate and sulphide minerals[38,45]. The thickness of the solid line reflects 1 s.d. of the average $\Delta_{\text{pyrite-seawater}}$ from ref. 38. The green band indicates the range of calculated $\Delta_{\text{pyrite-seawater}}$ values for experiments where the pyrite burial flux was 1×10^{18} mol S Myr^{-1} and the δ^{34}S value of riverine was varied between 0 and 10‰. The blue band indicates the range of calculated $\Delta_{\text{pyrite-seawater}}$ values for experiments where the pyrite burial flux was 0.67×10^{18} mol S Myr^{-1} and the δ^{34}S value of riverine was varied between 0 and 10‰. The brown band indicates the range of calculated $\Delta_{\text{pyrite-seawater}}$ values for experiments where the pyrite burial flux was 0.22×10^{18} mol S Myr^{-1} and the δ^{34}S value of riverine was varied between 0 and 10‰.



Extended Data Figure 3 | Calculated $\delta^{34}S$ values of riverine input from 0 to 30 Myr ago for constant $\Delta_{pyrite\text{-}seawater}$ values and pyrite burial fluxes. The solid black line on the vertical axis indicates the range of estimates for the modern $\delta^{34}S$ value of riverine input[27,40]. The green band indicates the range of calculated $\delta^{34}S$ values of riverine input for experiments where the pyrite burial flux was 1×10^{18} mol S Myr^{-1} and $\Delta_{pyrite\text{-}seawater}$ was varied between −30 and −50‰. The blue band indicates the range of calculated $\delta^{34}S$ values of riverine input for experiments where the pyrite burial flux was 0.67×10^{18} mol S Myr^{-1} and $\Delta_{pyrite\text{-}seawater}$ was varied between −30 and −50‰. The brown band indicates the range of calculated $\delta^{34}S$ values of riverine input for experiments where the pyrite burial flux was 0.22×10^{18} mol S Myr^{-1} and $\Delta_{pyrite\text{-}seawater}$ was varied between −30 and −50‰.



Extended Data Figure 4 | Integrated O_2 consumption and CO_2 production implied by the S cycle inverse model for different input parameters. The O_2 consumption values are presented in molar units. Importantly, the implied p_{O_2} values ignore other processes affecting the O_2 budget (for example changes in the organic C cycle). Negative values of both CO_2 and O_2 consumption reflect net sulphide burial. Independent estimates of the $\delta^{34}S$ of riverine input[40,46] and the average $\Delta_{\text{pyrite-seawater}}$ (ref. 38) correspond to the parameter combinations that produce net sulphide oxidation (that is, positive values) and are marked with stars. **a**, The model results integrated from 0 to 50 Myr ago. **b**, The model results integrated from 0 to 15 Myr ago following the results of the Li and Elderfield[10] model, which suggests that sulphide oxidation acted at least in part to balance enhanced CO_2 consumption over this time period. The dashed horizontal lines represent the CO_2 release required to compensate fully for a 10% (blue), 25% (purple) or 50% (orange) increase in continental silicate weathering that occurred linearly over the past 15 Myr assuming a modern flux of 3.5×10^{18} mol Myr^{-1} (Methods). The 25% increase corresponds to the increase predicted by the Li and Elderfield[10] mass balance model.

Extended Data Table 1 | CO_2 consumption and release rates in modern river systems[17,19–22]

mol/yr/km^2	Na^+ silicate	K^+ silicate	Mg^{2+} silicate	Ca^{2+} silicate	CO_2 consumption (uncertainty)
Liwu	2.6E+05	1.7E+05	6.3E+04	9.3E+04	2.1E+05 (6.7E+04)
GB	6.8E+04	2.8E+04	1.4E+04	1.4E+04	4.0E+04 (1.6E+04)
Mackenzie	2.1E+04	3.5E+03	5.3E+03	6.8E+03	1.6E+04 (4.7E+03)
	Ca^{2+} carbonate	**Mg^{2+} carbonate**	**SO_4^{2-} total**	**SO_4^{2-} sulfide**	**CO_2 release (uncertainty)**
Liwu	3.2E+06	9.1E+05	2.1E+06	1.8E+06	1.8E+06 (1.3E+05)
GB	2.3E+05	8.9E+04	7.9E+04	5.5E+04	5.5E+04
Mackenzie	1.4E+05	6.4E+04	7.6E+04	6.2E+04	6.2E+04 (3.1E+03)

doi:10.1038/nature12973

Melanosome evolution indicates a key physiological shift within feathered dinosaurs

Quanguo Li[1], Julia A. Clarke[2], Ke-Qin Gao[3], Chang-Fu Zhou[4], Qingjin Meng[5], Daliang Li[6], Liliana D'Alba[7] & Matthew D. Shawkey[7]

Inference of colour patterning in extinct dinosaurs[1–3] has been based on the relationship between the morphology of melanin-containing organelles (melanosomes) and colour in extant bird feathers. When this relationship evolved relative to the origin of feathers and other novel integumentary structures, such as hair and filamentous body covering in extinct archosaurs, has not been evaluated. Here we sample melanosomes from the integument of 181 extant amniote taxa and 13 lizard, turtle, dinosaur and pterosaur fossils from the Upper-Jurassic and Lower-Cretaceous of China. We find that in the lineage leading to birds, the observed increase in the diversity of melanosome morphologies appears abruptly, near the origin of pinnate feathers in maniraptoran dinosaurs. Similarly, mammals show an increased diversity of melanosome form compared to all ectothermic amniotes. In these two clades, mammals and maniraptoran dinosaurs including birds, melanosome form and colour are linked and colour reconstruction may be possible. By contrast, melanosomes in lizard, turtle and crocodilian skin, as well as the archosaurian filamentous body coverings (dinosaur 'protofeathers' and pterosaur 'pycnofibres'), show a limited diversity of form that is uncorrelated with colour in extant taxa. These patterns may be explained by convergent changes in the key melanocortin system of mammals and birds, which is known to affect pleiotropically both melanin-based colouration and energetic processes such as metabolic rate in vertebrates[4], and may therefore support a significant physiological shift in maniraptoran dinosaurs.

Melanin-based colour is a ubiquitous feature of amniote integument, found in feathers of every major bird clade[5,6], as well as in skin and hair (the scales covering skin in reptiles are typically transparent[7]). We examined the relationship between integumentary structure and melanosome morphology in a phylogenetic context. Individual melanosomes were measured from scanning electron microscope (SEM) images of the hairs of 44 species of extant mammals ($n = 51$ samples) and the skin of 36 extant species ($n = 36$ samples), sampling across Lepidosauria, Testudines and Crocodylia (Supplementary Tables 1–3). These measurements were compared with our previously published data from 101 extant avian species ($n = 168$ feather samples[3]). Samples targeting the full range of melanin-based colours for each integument type included blacks, browns and greys for feathers, hair and skin. As feathers uniquely show a broad range of melanin-based iridescent colours, and these are known to be associated with distinctive melanosome morphologies[3], they were also included. Although iridescent colours are found in reptile skin[8] and (rarely) in hair[9], they are produced through light scattering from iridiphores and multilayer keratin films, respectively. As these colours are not melanin-based, the integuments in which they are produced were not sampled. To capture melanosome diversity fully also demanded inclusion of morphotypes associated with black colour that are so far known only from extant penguins[10].

Preserved integument from 13 fossil amniotes from the Upper Jurassic and Lower Cretaceous of northeast China were sampled using described protocols[2,3,10] (Supplementary Methods). Skin was sampled from two fossil lepidosaurs, a turtle, and two specimens of the ornithischian dinosaur *Psittacosaurus* (Extended Data Figs 1–3 and Supplementary Tables 2 and 3). Although extant amniote epidermal appendages are limited to scales, hairs and feathers[11], additional structures are observed in the fossil record. These include basally bunched or single hollow filaments in theropod dinosaurs proposed to be homologous with modern feathers[12–14], bristles and filaments in ornithischian dinosaurs[15,16], and pterosaur pycnofibres, the identity and homology of which are more controversial[17]. Filamentous structures were sampled from the theropod dinosaur *Beipiaosaurus* and two pterosaurs (Extended Data Figs 4, 5 and Supplementary Tables 2, 3). Feathers were sampled from *Caudipteryx*, *Confuciusornis*, one ornithurine and two enantiornithine birds (Extended Data Figs 2, 6–8 and Supplementary Tables 2 and 3). Published SEM images and data were assessed for filamentous structures in the non-maniraptoran coelurosaurian dinosaur *Sinosauropteryx*[1], feathers from the basal paravians *Anchiornis*[2], *Microraptor*[3] and *Archaeopteryx*[18] and

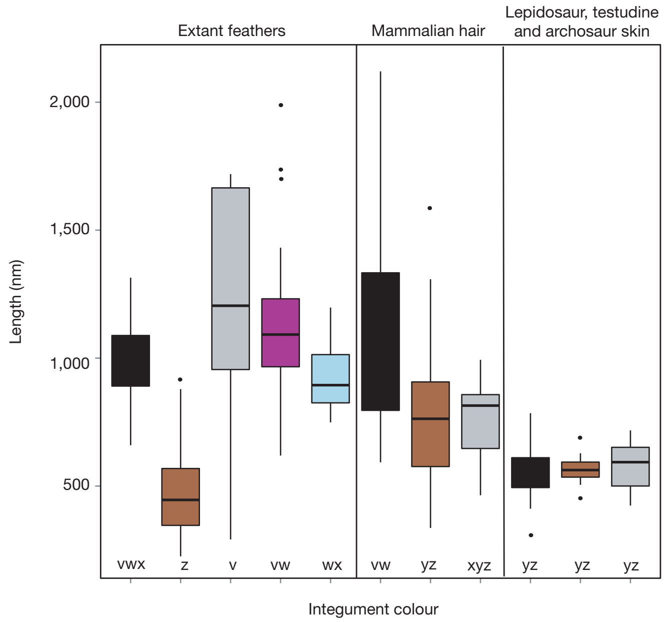


Figure 1 | Melanosome length observed in extant feathers, lepidosaur, testudine and archosaur skin, and mammalian hair. Boxplot colours correspond with integument colour: black, brown and grey. For feathers, 'penguin-like' is shown in blue, and iridescent is shown in purple. Lines are median values, boxes are quartiles, lines are range. Boxplots sharing the same letter (v, w, x, y, z) are not significantly different (two-sided Tukey HSD; $P < 0.05$) from one another; melanosome shape correlates with distinct colours in feathers and hair but not in skin. Extant feathers, $n = 168$; lepidosaur, testudine and archosaur skin, $n = 36$; mammalian hair, $n = 51$.

[1]State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Beijing 100083, China. [2]Department of Geological Sciences, University of Texas at Austin, 1 University Station C1100, Austin, Texas 78712, USA. [3]School of Earth and Space Sciences, Peking University, Beijing 100871, China. [4]Institute of Paleontology, Shenyang Normal University, Shenyang 110034, China. [5]Beijing Museum of Natural History, 126 Tianqiao South Street, Beijing 100050, China. [6]Museum of China University of Geosciences (Beijing), 29 Xueyuan Road, 100083, China. [7]Department of Biology and Integrated Bioscience Program, University of Akron, Akron, Ohio 44325-3908, USA.

the extinct penguin *Inkayacu*[10]. We used factorial analyses of variance (ANOVAs) and Tukey HSD (honestly significant difference) post-hoc tests in the R programming language[19] to test for differences in melanosome morphology by colour and integument type in extant samples (Fig. 1, Supplementary Methods and Extended Data Fig. 9). Canonical discriminant analysis (CDA) assessed the accuracy of the colours predicted for extant skin and hair samples based on melansome morphology (Supplementary Methods).

Across extant Amniota, the range of melanosome shape and form differs between taxa and integumentary types (Figs 1–4). These differences are directly related to the ability to predict colour from melanosome morphology (Supplementary Methods). For example, unlike in extant bird feathers[1–3], extant lepidosaur, turtle and crocodilian skin colour cannot be predicted from morphology alone (Fig. 1; Supplementary Methods). Diversity of melanosome morphology in these taxa is limited (Figs 1 and 2a, b) compared to mammals and birds (Figs 2c–f and 4). Although CDA analyses of extant feathers classify colours with 82% accuracy based on morphology[3], no two colours of skin could be discriminated. Indeed, because no morphological variable was correlated with colour in extant skin, none could be used in CDA (Supplementary methods). Overall, melanosome length was more variable than diameter within and among amniote integument types. For hair (Fig. 2c, d), a CDA with melanosome length as a variable weakly predicted colour when grey hairs were included (59% accuracy), but predicted colour well (87% accuracy) when they were excluded (Supplementary Methods). This is likely because, in contrast to grey feathers, grey hair is typically produced by macroscale patterns of alternating dark and light hairs or striping within single hairs, rather than associated with distinct melanosome shapes[7] (Supplementary Methods).

Only two shifts in integumentary structure in amniotes, the origin of pinnate feathers and mammalian hairs, are associated with shifts in melanosome diversity (Fig. 4). CDA indicates that colour and melanosome morphology are linked in the extant parts of these clades (Supplementary Methods). By contrast, no similar shift is observed in the transition to filamentous integument in archosaurs or in the transition from single hollow filaments to branched and basally joined bundles of filaments in Coelurosauria[15] (Fig. 4). If the structures in pterosaurs, ornithischian dinosaurs and theropods represent three separate origins of filamentous structures instead of one early in Archosauria, this pattern

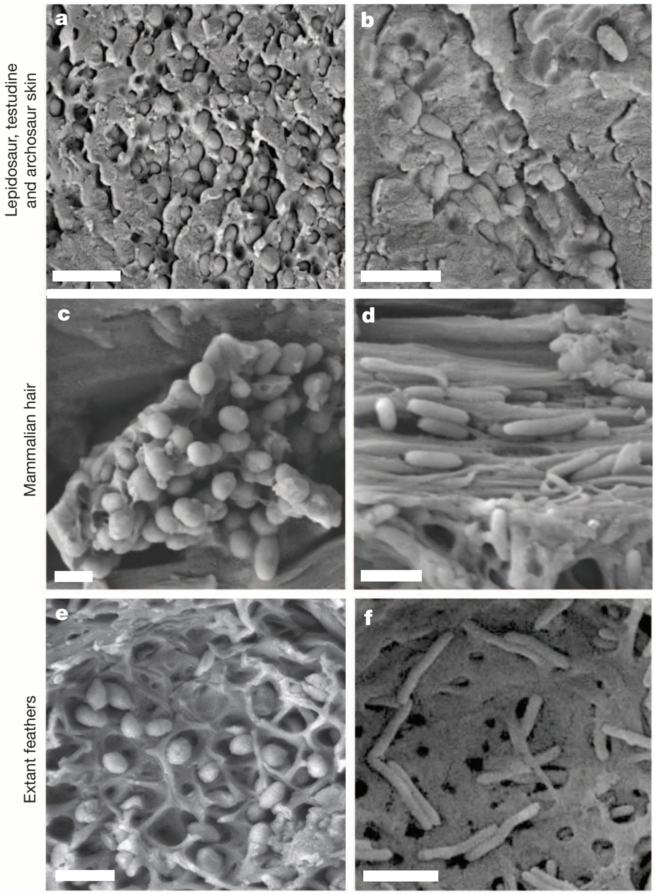


Figure 2 | SEM images of melanosomes from extant taxa representing the lowest-aspect-ratio and highest-aspect-ratio forms for each integumentary type. **a**, **b**, Skin of leaf-tailed gecko, *Uroplatus fantasticus* (**a**), and Galapagos fire lizard, *Microlophus albemarlensis* (**b**). **c**, **d**, Hair from hairy-tailed mole, *Parascalops breweri* (**c**), and domestic cat, *Felis catus* (**d**). **e**, **f**, Feathers from tufted titmouse, *Baeolophus bicolor* (**e**), and Huon astrapia, *Astrapia rothschildi* (**f**). Scale bars, 1 μm.

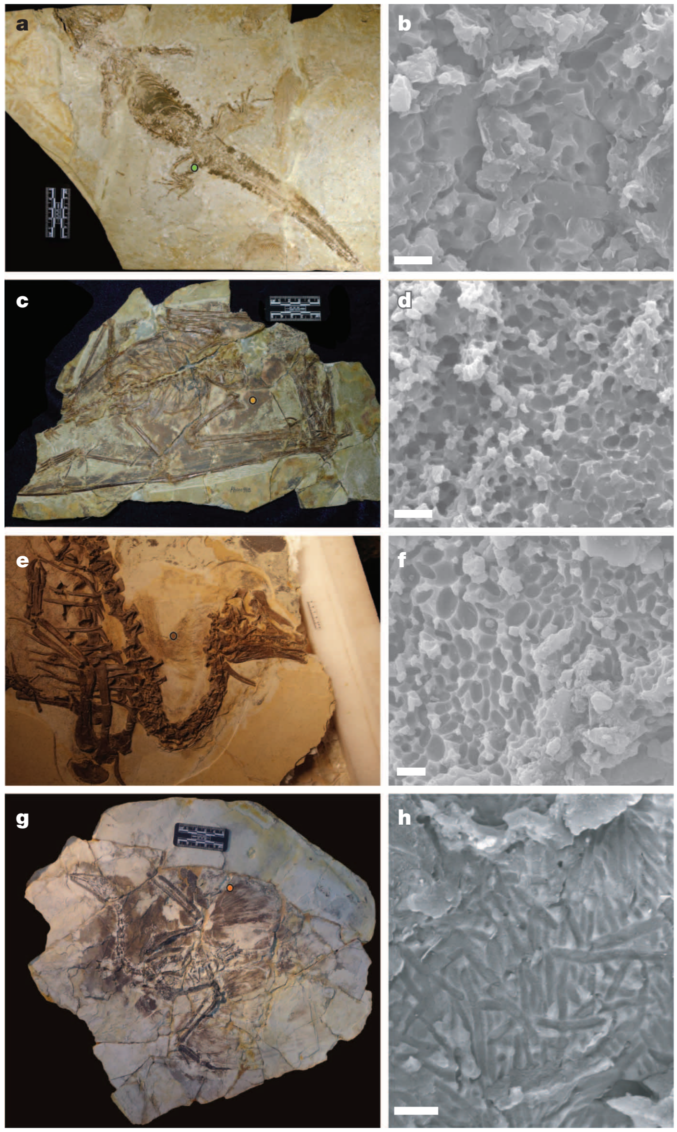


Figure 3 | Melanosomes from Jehol Group fossils. **a**–**h**, Images show the differences between low-aspect melanosomes consistently observed in lepidosaur skin (PKUP V1059; **a**, **b**) and filamentous structures in pterosaurs (BMNHC PH000988; **c**, **d**); and between non-maniraptoran dinosaurs (*Beipiaosaurus*; BMNHC PH000911; **e**, **f**) and the high-aspect-ratio forms observed in maniraptoran feathers (undescribed ornithurine bird, CUGB G20100053; **g**, **h**). Dots in **a**, **c**, **e** and **g** indicate the location of the samples shown in **b**, **d**, **f** and **h**. Dot colours correspond to the colours in Fig. 4: light green, extinct lepidosaurian, testudine and crocodilian skin; dark brown, filamenous body covering in dinosaurs; light brown, pterosaurs; orange, feathers in extinct stem avialan taxa. See Extended Data Figures for complete sampling maps. Scale bars, 1 μm.

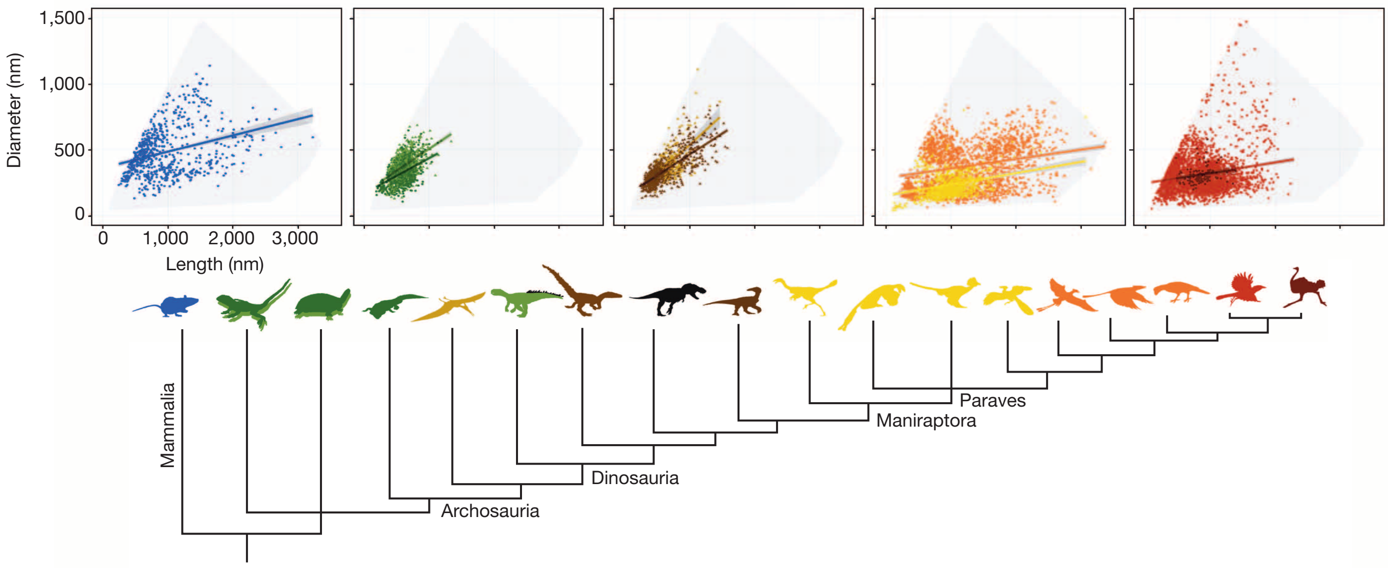


Figure 4 | Melanosome diversity across Amniota. Scatterplots of individual melanosome measurements (diameters and lengths) by integument type for extant mammal hair (blue, $n = 719$), skin from extant (dark green, $n = 742$) and extinct (light green, $n = 925$) lepidosaurian, testudine and archosaurian species, filamentous body covering in pterosaurs (light brown, $n = 899$), filamentous body covering in dinosaurs (dark brown, $n = 164$), feathers in basal Paraves (yellow, $n = 1{,}268$), *Confuciusornis* and crown-ward extinct avialan taxa (orange, $n = 1{,}683$), extant Aves (bright red, $n = 3{,}294$) and flightless palaeognath birds (dark red, $n = 107$). Colours of silhouettes correspond with colours in scatterplots. Black indicated unsampled taxa or integumentary type (for example, bristle structures on the tail of *Psittacosaurus*). Grey regions in scatterplots indicate the extent of melanosome diversity in all samples (total observed melanosome morphospace), and regression lines are drawn for comparison among integumentary types. Shaded areas around lines indicate 95% confidence intervals.

would be even stronger. Observed melanosome size and variation were similar in extinct and extant taxa of each major amniote group and for each integumentary type (for example, extant and fossil skin) (Figs 2–4). Accounting for proposed taphonomic alteration of fossil melanosome size (that is, average 18–20% reduction in size[20]) does not affect recovered patterns (Extended Data Fig. 10; compare with Fig. 4). Indeed, fit of sampled fossil lepidosaur, testudine and archosaur melanosomes with extant taxa argues against significant taphonomic alteration (see also Supplementary Methods).

Pterosaurs and non-maniraptoran dinosaurs show a limited range of low-aspect-ratio (length:width, <2) melanosome morphologies (Figs 3c–f and 4) similar to the morphospace occupied by fossil and extant lepidosaurs, crocodylians and turtles (Figs 3a, b and 4). Low-aspect-ratio melanosomes in *Sinosauropteryx* filaments have been interpreted as indicative of a reddish brown colour based on avian comparisons[1]. By this logic the newly sampled pterosaurs and *Beipiaosaurus* (Fig. 3c, f and Extended Data Figs 4 and 5) would all be inferred as similarly brown (avian CDA; Supplementary Methods). However, plesiomorphically, low-aspect-ratio melanosomes from skin (Fig. 2a, b) produce black, brown or grey colours, and there is no evidence that derived avian relationships between shape and colour apply to more basal taxa occupying a similar (outgroup) melanosome morphospace (Fig. 1, Extended Data Fig. 9 and Supplementary Table 2). Colour reconstruction for these archosaurian taxa based on melanosome morphology is therefore cautioned against.

Melanosome diversity in Amniota does not track novelties in the material of which these integumentary structures are comprised, previously proposed differences in melanin chemistry, or the range of colours produced (Fig. 4). In terms of novelties in material, novel β-keratins arose in both Lepidosauria and Testudines, some are shared by birds and crocodilians, and some are unique to birds[21]. However, only those unique to birds may be associated with observed changes in the melanin system. Mammalian hair consists of structurally distinct α-keratin but exhibits melanosome diversity approaching that in birds (Figs 2 and 4). In terms of previously proposed differences in melanin chemistry, although phaeomelanin was long thought to be exclusive to mammals and birds[22], it has recently been reported in amphibians[23] and turtles[24], and thus its distribution does not explain the recovered differences in melanosome diversity. Finally, in terms of the range of colours produced, colouration is a key function of melanin in amniotes and the range of colours across our sampled taxa vary markedly. However, melanosome diversity in amniotes is not predicted by how colourful the extant parts of these taxa are today. Lepidosaurs show low melanosome diversity and brightly coloured species make use of other colour-producing mechanisms[8]. Mammals generally exhibit more muted colours than those observed in birds and lack structural colouration other than a weak iridescence produced by a keratin multilayer in golden moles[9]. However, they show a striking diversity of melanosome morphologies, most similar to the range observed in birds (Fig. 4).

Shifts in the melanin-based colour system in amniotes may instead be explained by physiological innovations. Convergent changes in the melanin-based colour system towards an expanded range of melanosome lengths, diameters and aspect ratios occur in maniraptoran dinosaurs and in mammals (Figs 2–4). Interactions between the agouti signalling protein (ASIP) and the melanocortin 1 receptor (MC1R) have been implicated in shifts in melanin synthesis in developing feathers[25] as well as hair[26]. Within-feather patterning, such as stripes, results from these interactions[25,26] and first appears within maniraptoran dinosaurs[1] coincident with the increase in melanosome diversity noted here (Fig. 4). The changes we observe are thus consistent with independent shifts in at least these axes of the melanocortin system in the lineages leading to mammals and birds.

The melanocortin system pleiotropically affects both melanin-based colouration as well as key energetic processes such as metabolic rate, reproductive physiology, stress axis and food intake across vertebrates (from bony fish to humans[4,27,28]). In Aves, inter- and intraspecific variation in melanic colour has been linked to differences in energetics and metabolism[4,27], and a possible role of convertase enzymes PC1/3 and PC2 was recently implicated in pleiotropic effects[28]. ASIP also has well-characterized pleiotropic effects[27], and in particular the ASIP–MC1R interaction was recently recognized to have a physiological as well as pigmentary role in birds[29]. Mirroring the broad patterns between ecto- and endothermic animals, our results show that avian taxa that exhibit

lower basal metabolic rates, flightless paleognaths[30], also exhibit lower melanosome diversity (Fig. 4). The fact that melanosome diversity in endothermic extant amniotes is similar, despite profound differences in keratin chemistry[21], integumentary structure and melanocortin pathways[4,27], may be because melanosome diversity is pleiotropically linked to changes in energetics associated with the higher metabolic rates they uniquely share. Although further investigation is clearly required, the change in melanin-based colour reported here, and the origin of within-feather melanin-based colouration in Maniraptora[1], may indicate when a key shift in dinosaurian physiology occurred before the origin of flight.

METHODS SUMMARY

We obtained data on melanosome morphology from extant hair and skin samples using techniques previously described for feathers[3]. In brief, hair and skin samples were embedded in plastic and sectioned longitudinally with a glass knife on an ultramicrotome. These sections were viewed on a JEOL JSM7401F SEM. The image-processing program ImageJ (available for download at http://rsbweb.nih.gov/ij/) was used to measure length, diameter, aspect ratio and density in SEM images acquired. We used ANOVA and Tukey post-hoc tests to compare these variables between integumentary type, and a canonical discriminant analysis to determine whether colour of skin or hair could be predicted from these morphological variables. We took small (approximately 1 mm^2) samples from a range of integumentary structures in the fossil record, including feathers, skin, 'pycnofibres' and 'protofeathers' and viewed them on a ZEISS SUPRA-55 VP SEM. Melanosome morphology data were collected from these samples using ImageJ. We visualized all data using the *ggplot2* package in R[19].

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 October; accepted 20 December 2013.
Published online 12 February 2014.

1. Zhang, F. *et al.* Fossilized melanosomes and the colour of Cretaceous dinosaurs and birds. *Nature* **463,** 1075–1078 (2010).
2. Li, Q. *et al.* Plumage color patterns of an extinct dinosaur. *Science* **327,** 1369–1372 (2010).
3. Li, Q. *et al.* Reconstruction of *Microraptor* and the evolution of iridescent plumage. *Science* **335,** 1215–1219 (2012).
4. Ducrest, A.-L., Keller, L. & Roulin, A. Pleiotropy in the melanocortin system, coloration and behavioural syndromes. *Trends Ecol. Evol.* **23,** 502–510 (2008).
5. Stoddard, M. C. & Prum, R. O. How colorful are birds? Evolution of the avian plumage color gamut. *Behav. Ecol.* **22,** 1042–1052 (2011).
6. McGraw, K. J. in *Bird Coloration. 1. Mechanisms and Measurements* (eds Hill, G. E. & McGraw, K. J.) 243–294 (Harvard Univ. Press, 2006).
7. Mills, M. G. & Patterson, L. B. Not just black and white: pigment pattern development and evolution in vertebrates. *Semin. Cell Dev. Biol.* **20,** 72–81 (2009).
8. Bagnara, J. T. & Hadley, M. E. *Chromatophores and Color Change: The Comparative Physiology of Animal Pigmentation* (Prentice-Hall, 1973).
9. Snyder, H. K. *et al.* Iridescent colour production in hairs of blind golden moles (Chrysochloridae). *Biol. Lett.* **8,** 393–396 (2012).
10. Clarke, J. A. *et al.* Fossil evidence for evolution of the shape and color of penguin feathers. *Science* **330,** 954–957 (2010).
11. Spearman, R. I. C. *The Integument* (Cambridge Univ. Press, 1973).
12. Xu, X., Zhou, Z. & Prum, R. O. Branched integumental structures in *Sinornithosaurus* and the origin of feathers. *Nature* **410,** 200–204 (2001).
13. Xu, X. *et al.* Basal tyrannosauroids from China and evidence for protofeathers in tyrannosauroids. *Nature* **431,** 680–684 (2004).
14. Ji, Q., Norell, M. A., Gao, K.-Q., Ji, S.-A. & Ren, D. The distribution of integumentary structures in a feathered dinosaur. *Nature* **410,** 1084–1088 (2001).
15. Zheng, X.-T., You, H.-L., Xu, X. & Dong, Z.-M. An Early Cretaceous heterodontosaurid dinosaur with filamentous integumentary structures. *Nature* **458,** 333–336 (2009).
16. Mayr, G., Peters, D. S., Plodowski, G. & Vogel, O. Bristle-like integumentary structures at the tail of the horned dinosaur *Psittacosaurus. Naturwissenschaften* **89,** 361–365 (2002).
17. Kellner, A. W. A. *et al.* The soft tissue of *Jeholopterus* (Pterosauria, Anurognathidae, Batrachognathinae) and the structure of the pterosaur wing membrane. *Proc. R. Soc. Lond. B* **277,** 321–329 (2010).
18. Carney, R., Vinther, J., Shawkey, M. D., D'Alba, L. & Ackermann, J. New evidence on the colour and nature of the isolated *Archaeopteryx* feather. *Nature Comm.* **3,** 637 (2012).
19. R Development Core Team. R: A language and environment for statistical computing. (R Foundation for Statistical Computing, 2008).
20. McNamara, M. E., Briggs, D. E. G., Orr, P. J., Field, D. J. & Wang, Z. Experimental maturation of feathers: implications for reconstructions of fossil feather colour. *Biol. Lett.* **9,** 20130184.
21. Greenwold, M. H. & Sawyer, R. H. Linking the molecular evolution of avian beta (β) keratins to the evolution of feathers. *J. Exp. Zool. Mol. Dev. Evol.* **316,** 609–616 (2011).
22. Ito, S. & Wakamatsu, K. Quantitative analysis of eumelanin and pheomelanin in humans, mice, and other animals. *Pigment Cell Res.* **16,** 523–531 (2003).
23. Wolnicka-Glubisz, A., Peclo, A., Podkowa, D., Kolodziejczyk, L. M. & Plonka, P. M. Pheomelanin in the skin of *Hymenochirus boettgeri* (Amphibia:Anura:Pipidae). *Exp. Dermatol.* **21,** 537–540 (2012).
24. Roulin, A., Mafli, A. & Wakamatsu, K. Reptiles produce pheomelanin: evidence in the Eastern Hermann's Tortoise. *J. Herpetol.* **47,** 258–261 (2013).
25. Yoshihara, C. *et al.* Elaborate color patterns of individual chicken feathers may be formed by the agouti signaling protein. *Gen. Comp. Endocrinol.* **175,** 495–499 (2012).
26. Manceau, M., Domingues, V. S., Mallarino, R. & Hoekstra, H. E. The developmental role of agouti in color pattern evolution. *Science* **331,** 1062–1065 (2011).
27. Hubbard, J. K., Uy, J. A. C., Hauber, M. E., Hoekstra, H. E. & Safran, R. J. Vertebrate pigmentation: from underlying genes to adaptive function. *Trends Genet.* **26,** 231–239 (2010).
28. Emaresi, G. *et al.* Pleiotropy in the melanocortin system: expression levels of this system are associated with melanogenesis and pigmentation in the tawny owl (*Strix aluco*). *Mol. Ecol.* **22,** 4915–4930 (2013).
29. Yabuuchi, M., Bando, K., Hiramatsu, M., Takahashi, S. & Takeuchi, S. Local agouti signaling protein/melanocortin signaling system that possibly regulates lipid metabolism in adipose tissues of chickens. *J. Poult. Sci.* **47,** 176–182 (2010).
30. Calder, W. A., III & Dawson, T. J. Resting metabolic rates of ratite birds: the kiwis and the emu. *Biochem. Physiol. A* **60,** 479–481 (1978).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the National Natural Science Foundation of China (NSFC) grant 41272031, Fundamental Research Funds for Central Universities, Beijing Municipal Bureau of Human Resources, NSF grants EAR-1251895 and 1251922, Human Frontier Science Program (HFSP) grant RGY-0083, Air Force Office of Scientific Research (AFOSR) grant FA9550-13-1-0222, and the Jurassic Foundation. The Smithsonian Institution (J. F. Jacobs and A. Wynn) and San Diego Museum of Natural History (P. Unitt) provided extant samples. BMNHC PH000911 was photographed by M. Ellison.

Author Contributions J.A.C., K.-Q.G., Q.L. and M.D.S. (listed alphabetically) jointly conceived the study and participated in manuscript preparation. Data for extant taxa were collected by L.D. and M.D.S. Data from fossil taxa were collected by Q.L., M.D.S., J.A.C., K.-Q.G., C.-F.Z., L.D. and Q.M. Data collection from fossils was supervised by Q.L., Q.M, C.-F.Z. and D.L.; J.A.C. and M.D.S. developed the analytical approach and assessed results jointly with Q.L. and K.-Q.G.; M.D.S, L.D. and Q.L. analysed the data.

Author Information Specimens are permanently reposited at the public institutions indicated in the text and Supplementary Table 3; sampling is illustrated in the Extended Data Figures, and melanosome data are given in Supplementary Table 2 or have been made available previously[3]. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A.C. (Julia_Clarke@jsg.utexas.edu) or M.D.S. (shawkey@uakron.edu).

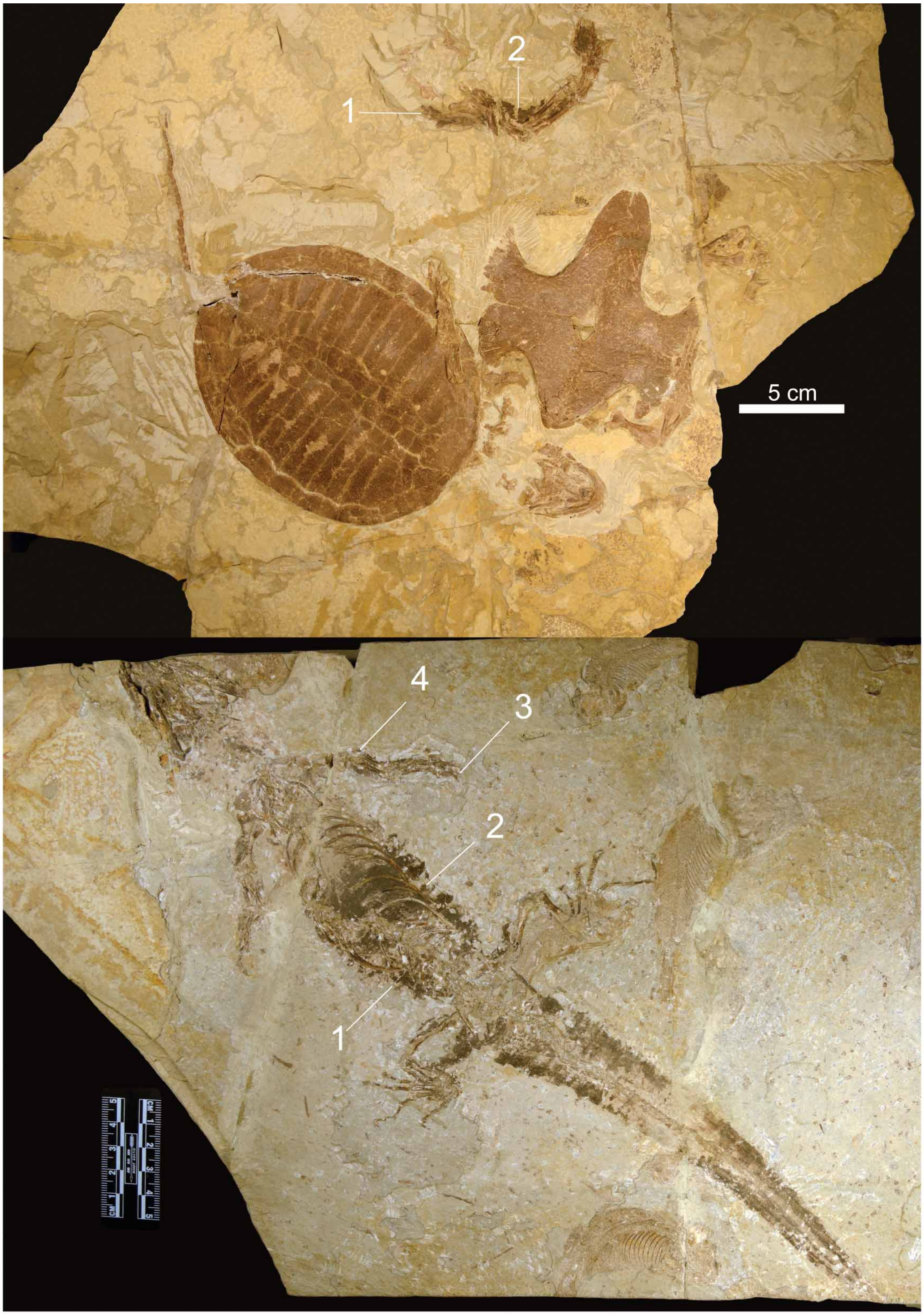
METHODS

Samples of extant amniote integument were used to capture the diversity of melanosomes associated with melanin-based colours found in each clade (see Supplementary Methods for details). Extant hairs were pulled from study specimens of mammals in the collections of the University of Akron and the San Diego Museum of Natural History, and small (approximately $5\,mm^2$) sections of skin were cut from melanized portions near the forelimb of formalin-preserved reptiles in the collection of the Smithsonian Museum of Natural History. Sampled species are listed in Supplementary Table 2. We obtained data on melanosome morphology from these samples as described previously[3]. First, skin or hair was embedded in Epon by dehydration using 100% ethanol (20 min) twice, and by infiltration with 15, 50, 70 and 100% Epon (24 h each step). Infiltrated samples were then placed in block moulds and polymerized at 60 °C for 16 h. We then cut thick (5 μm) longitudinal sections of blocks with a glass knife on a Leica UC-6 ultramicrotome, mounted them on stubs with carbon tape, sputter-coated them with silver and viewed them on a JEOL JSM7401F SEM.

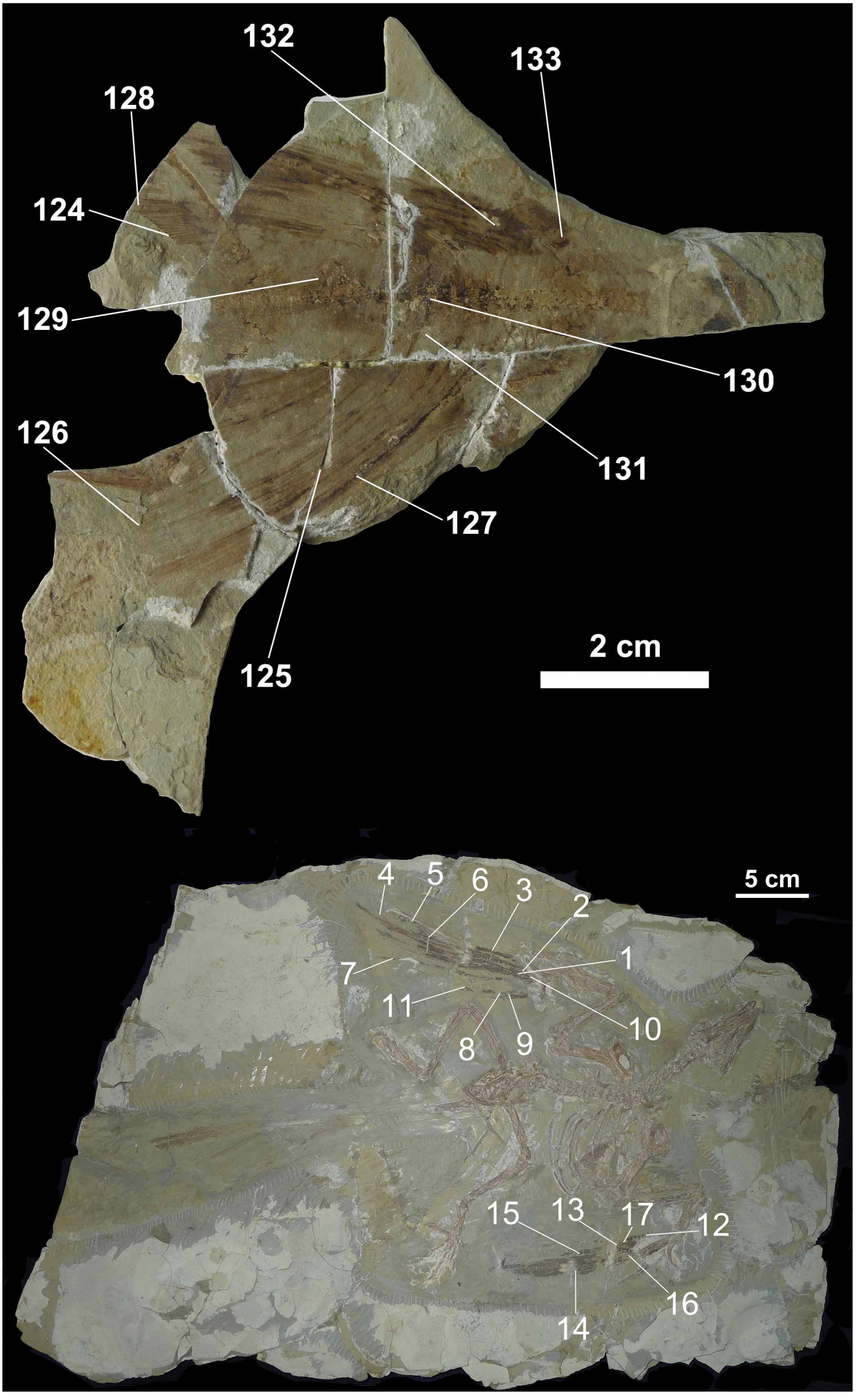
From the resulting SEM images we used the image-processing program ImageJ (available for download at http://rsbweb.nih.gov/ij/) to measure melanosome morphology as described previously[3]. We measured maximum linear short and long axis length of melanosomes that were oriented perpendicular to line of sight, and from these data calculated aspect ratio (long:short axis). These ratios are an index of shape, and values close to 1 indicate sphericity, whereas values further from 1 indicate cylindricity. The distribution of melanosome morphology within feathers was frequently skewed towards one type of morphology, therefore we also calculated the skewness of the long and short axes. We calculated density as the number of melanosomes per square micron of integumentary surface area. Using the R environment, we compared these values between extant coloured integument using ANOVA (aov function) and Tukey's post-hoc test (Tukey HSD function). Results are summarized in Fig. 1 and Supplementary Table 1.

We used a quadratic canonical discriminant analysis to determine whether colour could be predicted from melanososome morphology in hair and skin. We used a quadratic rather than linear discriminant analysis because some variables showed evidence of collinearity, but use of linear discriminant analysis did not significantly alter results. We used a forward stepwise method to choose variables that contributed significantly ($P < 0.05$) to the analysis and to eliminate those that did not.

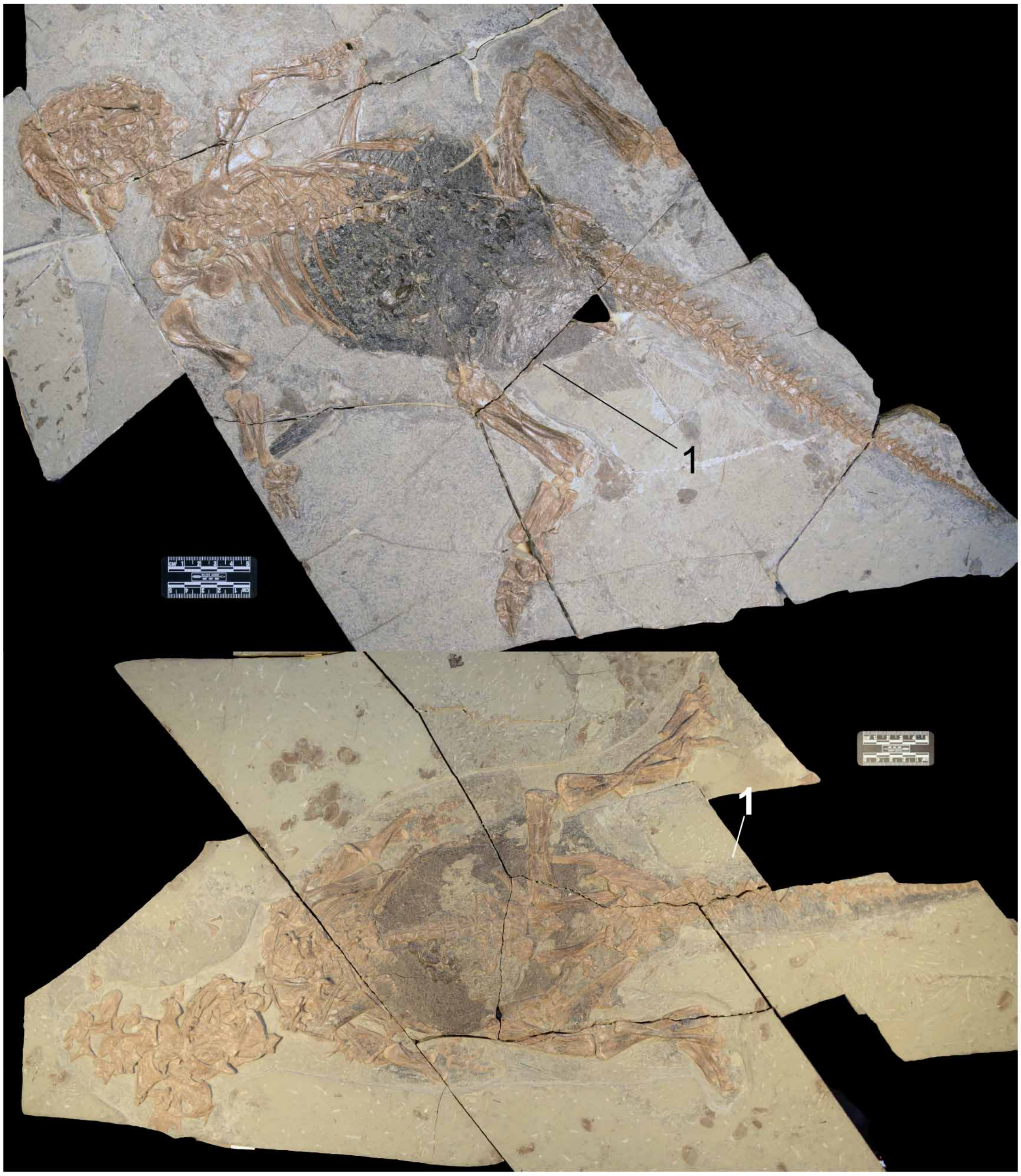
We sampled fossils with preserved integument spanning the range of known forms, including skin, 'protofeathers', 'pycnofibres' and pinnate feathers. We collected morphological data from SEM images of fossil integument from the Jehol biota in the same manner as the extant samples. Small (approximately $1\,mm^2$) samples were taken from preserved integument of fossils as before[3]. Each specimen was sampled as thoroughly as possible to maximize the potential diversity of melanosomes while maintaining the integrity of the fossil. The samples were coated with silver (30 s) and studied with either a ZEISS SUPRA-55 VP field emission scanning electron microscope (situated at China University of Geosciences, Beijing) or the JEOL environmental scanning electron microscope used for extant samples. Morphological measurements from melanosomes preserved as imprints and in three dimensions were assembled in the same manner as the modern samples. There was no taxonomic variation in the frequency of distinct preservational styles. These data are summarized in Fig. 3 and Supplementary Table 3.



Extended Data Figure 1 | Sampling map of an unnamed turtle fossil, PKUP V1070 (top) and a lizard fossil, *Yabeinosaurus sp.* PKUP V1059 (bottom). Numbers indicate integument sampling sites; melanosome data from all sites are presented in Fig. 4 and pooled for the per-taxon values given in Supplementary Table 2.

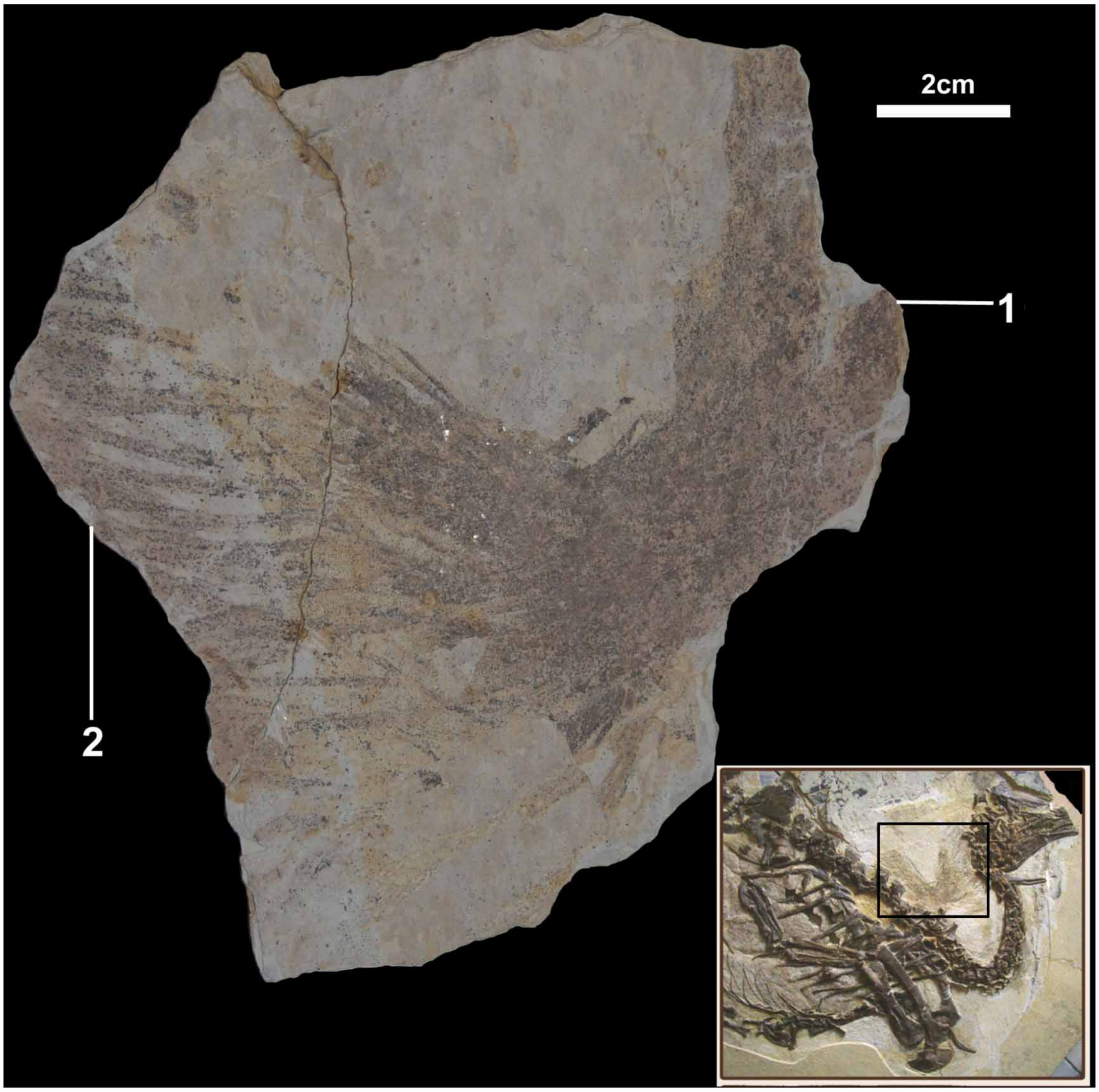


Extended Data Figure 2 | Sampling map of a gliding lizard fossil, *Xianglong zhaoi*, PMOL 000666, counterpart (top) and the basal avialan, *Confuciusornis sanctus*, CUGB G20070001 (bottom). Integument sampling sites are numbered. Numbers indicate integument sampling sites; melanosome data from all sites are presented in Fig. 4 and pooled for the per-taxon values given in Supplementary Table 2.

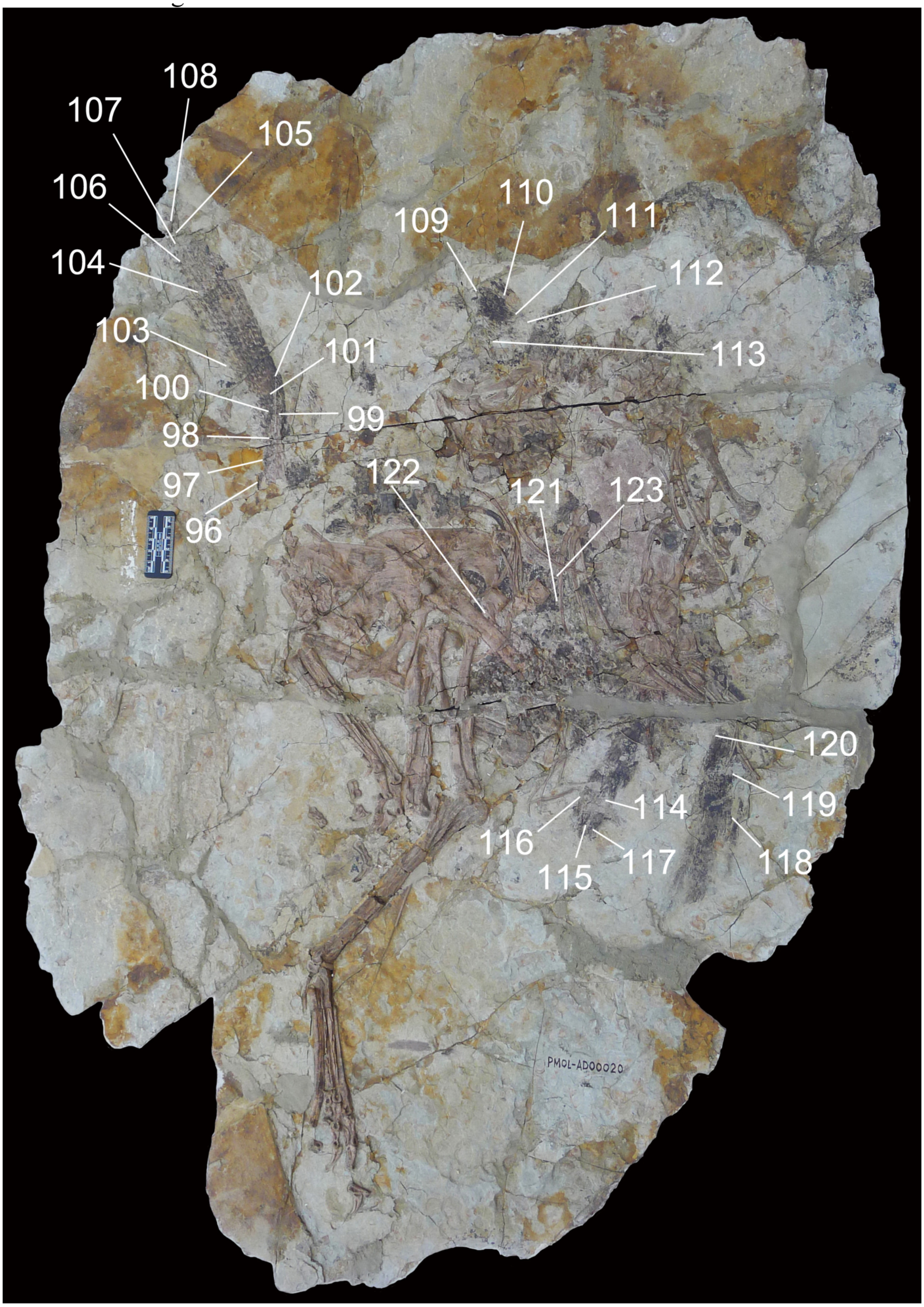


Extended Data Figure 3 | Skin sample from the ornithischian dinosaur, *Psittacosaurus lujiatunensis*, PKUP V1050 (top) and from the ornithischian dinosaur, *Psittacosaurus lujiatunensis*, PKUP V1051 (bottom). Numbers indicate integument sampling sites; melanosome data from all sites are presented in Fig. 4 and pooled for the per-taxon values given in Supplementary Table 2.

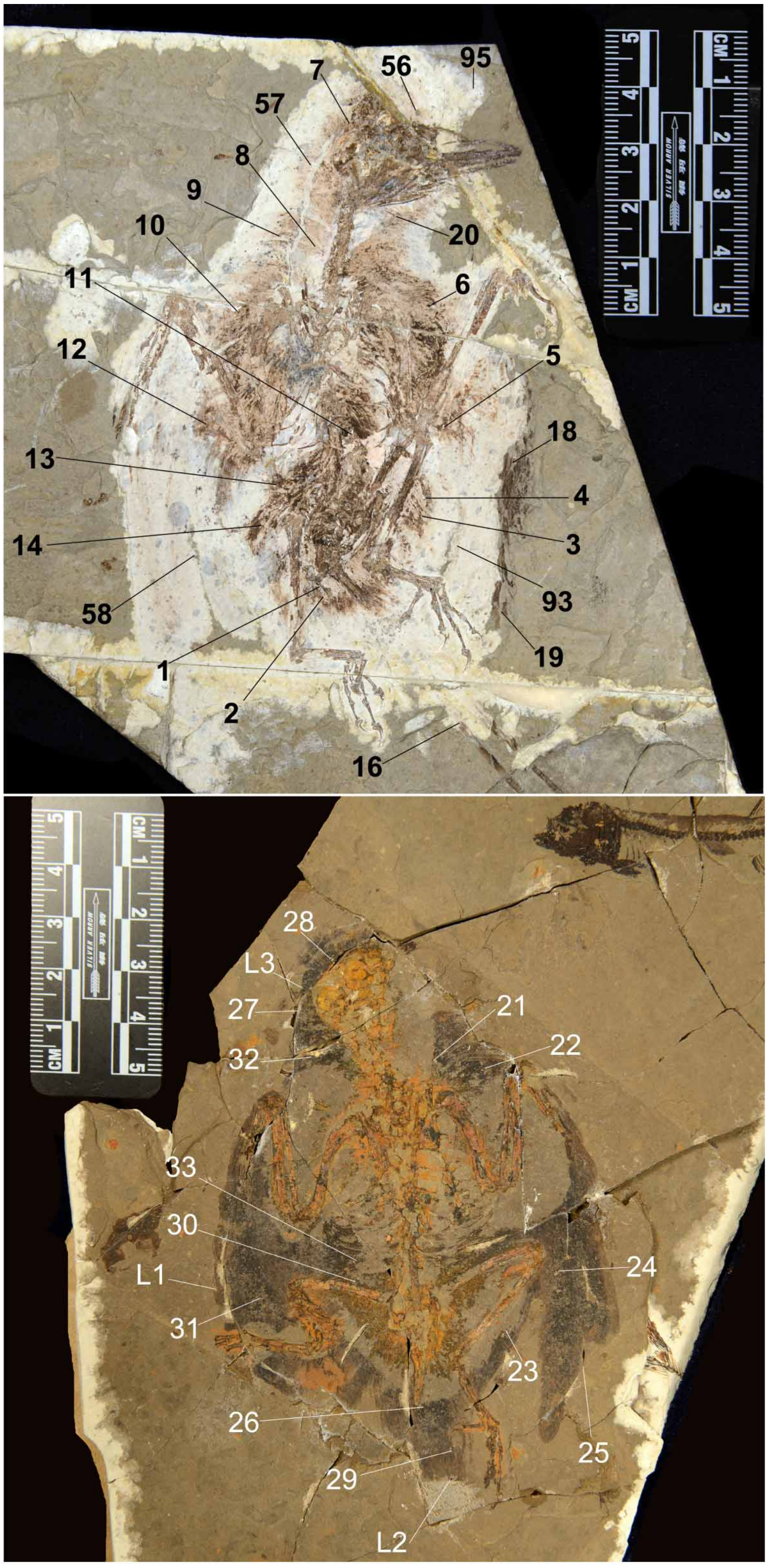
Extended Data Figure 4 | Sampling map for two pterosaurs, PMOL AP00022 (top) and BMNHC PH000988 (bottom). Numbers indicate integument sampling sites; melanosome data from all sites are presented in Fig. 4 and pooled for the per-taxon values given in Supplementary Table 2.



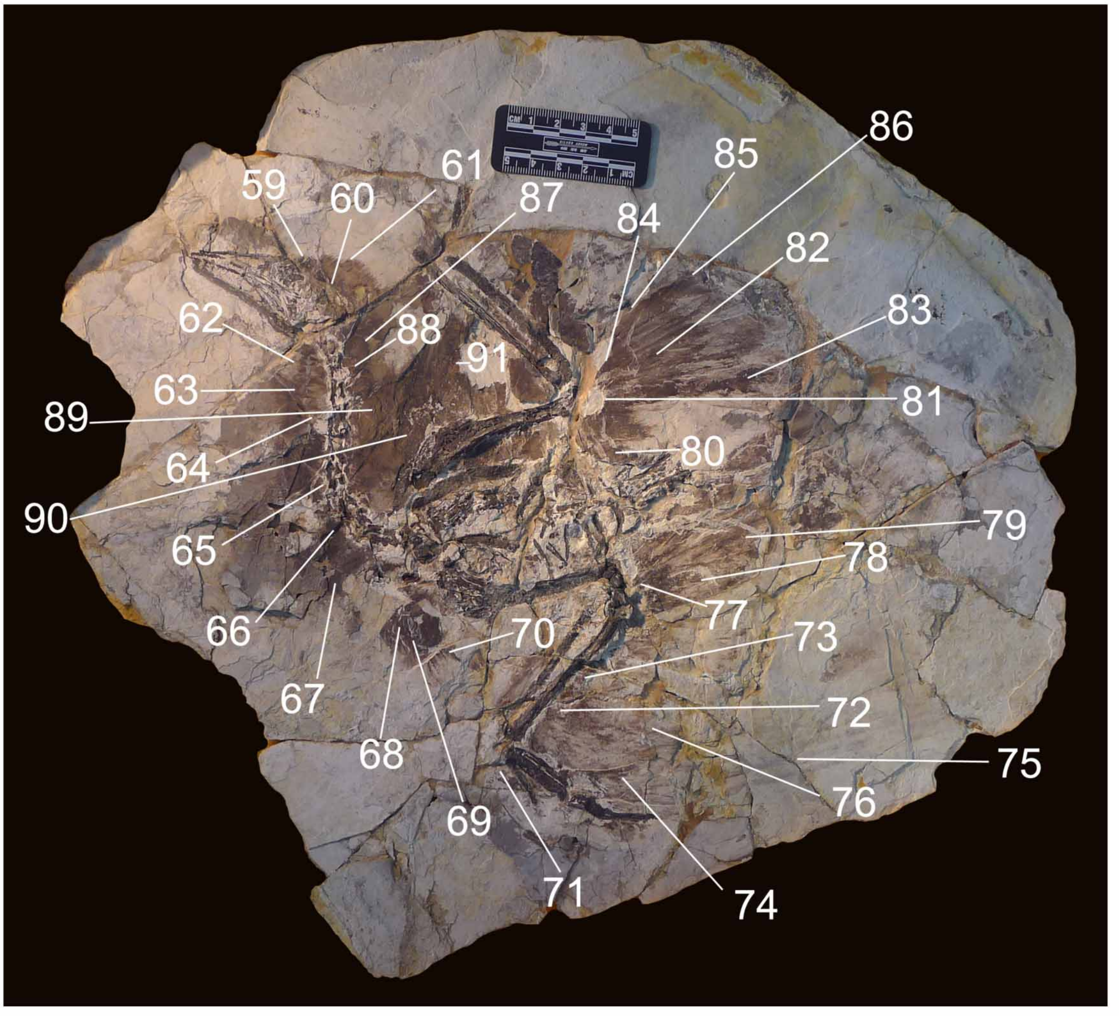
Extended Data Figure 5 | Samples from filaments preserved in the neck region of a skeleton of the theropod dinosaur, *Beipiaosaurus*, BMNHC PH000911 (counter slab). Numbers indicate integument sampling sites; melanosome data from all sites are presented in Fig. 4 and pooled for the per-taxon values given in Supplementary Table 2.



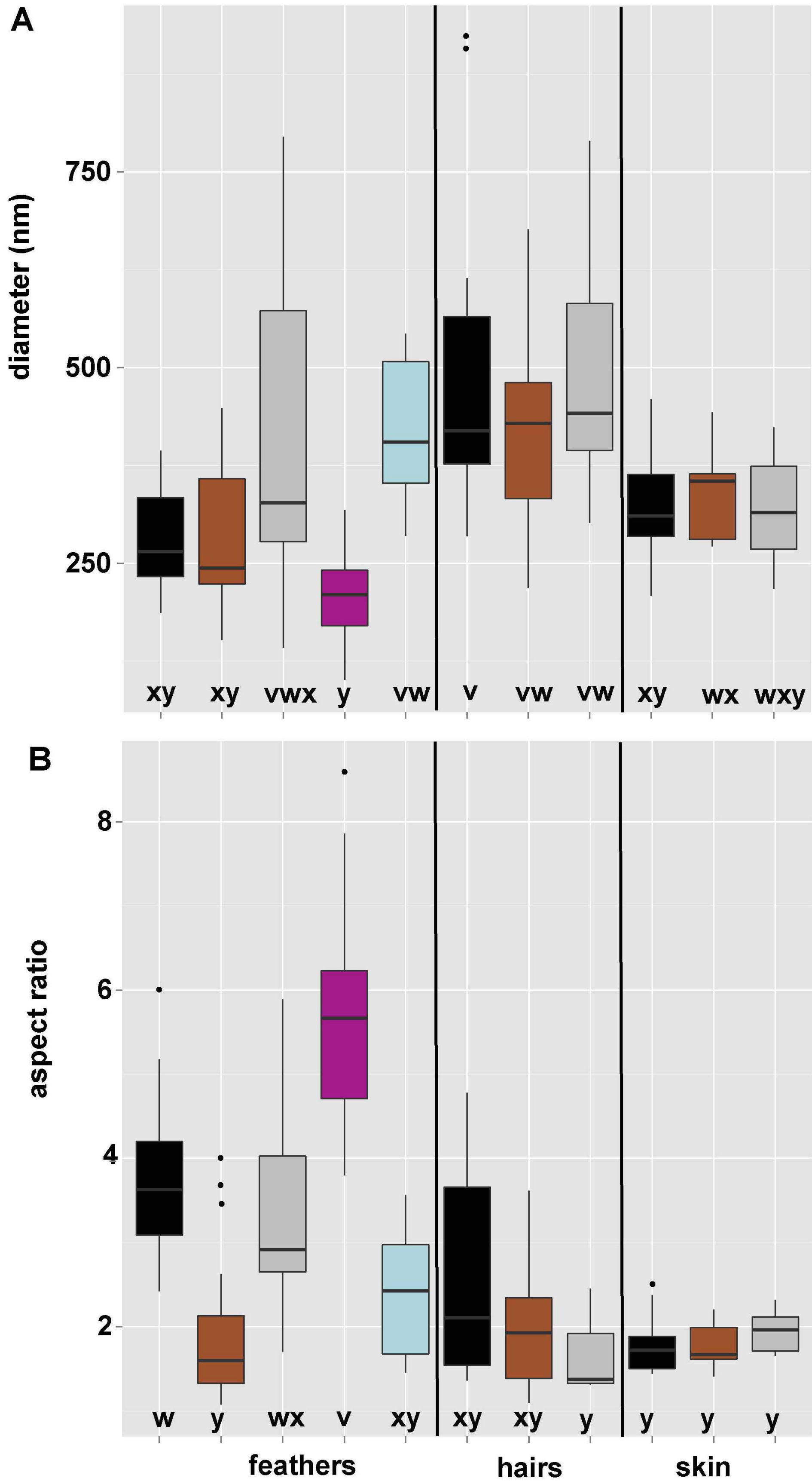
Extended Data Figure 6 | Sampling map of the feathered maniraptoran dinosaur, *Caudipteryx zoui*, PMOL AD00020. Numbers indicate integument sampling sites; melanosome data from all sites are presented in Fig. 4 and pooled for the per-taxon values given in Supplementary Table 2.



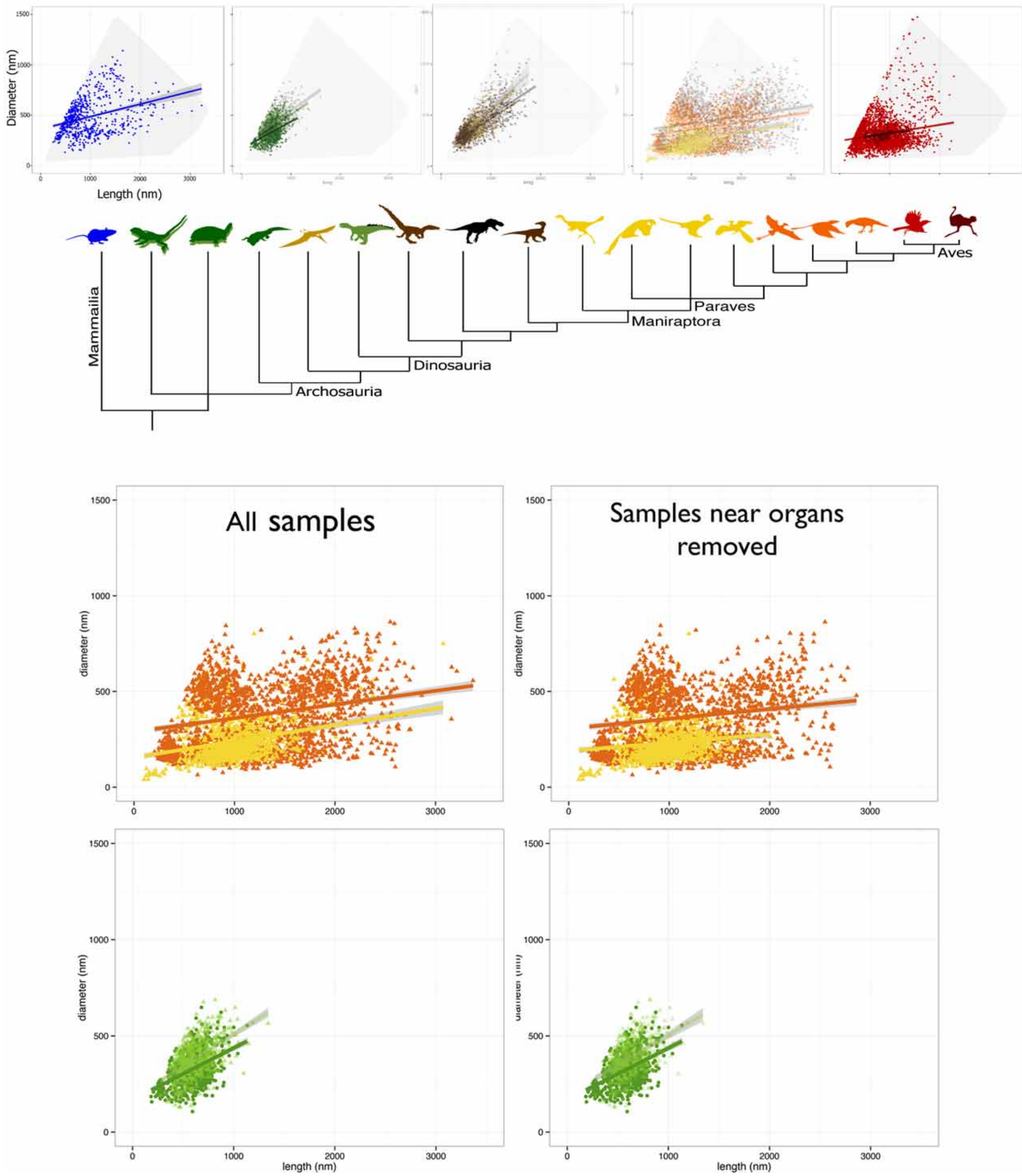
Extended Data Figure 7 | Sampling map of (top) an unnamed enantiornithine bird, CUGB G20120001 and (bottom) an undescribed enantiornithine bird, CUGB P1201. Numbers indicate integument sampling sites; melanosome data from all sites are presented in Fig. 4 and pooled for the per-taxon values given in Supplementary Table 2.



Extended Data Figure 8 | Sampling map of an undescribed ornithurine bird, CUGB G20100053. Numbers indicate integument sampling sites; melanosome data from all sites are presented in Fig. 4 and pooled for the per-taxon values given in Supplementary Table 2.



Extended Data Figure 9 | Melanosome diameters and aspect ratios observed in extant feathers, lepidosaur, testudine and archosaur skin, and mammalian hair. Melanosome diameters are shown in **a**, and aspect ratios are shown in **b**. Boxplot colours correspond with integument colour: black, brown, grey. For feathers, 'penguin-like' is shown in blue and iridescent is shown in purple. Lines are median values, boxes are quartiles, lines are range. Boxplots sharing the same letter (v, w, x, y, z) are not significantly different from one another.



Extended Data Figure 10 | Exploration of the potential effects of taphonomy and sampling on the observed differences in melanosomes in skin, hair, filaments and feathers. Top, melanosome diversity is adjusted to model taphonomic shrinkage of melanosomes suggested from experimental studies. Values for all fossil samples were adjusted (enlarged by 20%) based on the findings of ref. 18 (Supplementary Methods). Original data points are shown in colours and adjusted data are shown in grey. Grey regions indicate the extent of the total melanosome morphospace from the primary analyses. The pattern reported (i.e., increased diversity and higher-aspect-ratio forms only in Maniraptora and Mammalia) is not affected (Fig. 4, main text; *n* for each integumentary type is identical to the primary analysis). Bottom, to consider the effect of sampling on the observed pattern samples from near the thoracic region were removed from the database. Although there was no evidence to suggest that these samples were from internal organs, or that such organs were preserved, because melanosomes are present in some internal organs in extant taxa, the sensitivity of the results to removal these samples was explored. There was no effect on the pattern reported from the primary analysis (compare Fig. 4). Samples 1 and 2 from *Yabeinosaurus* sp. (PKUP V1059), sample 1 from *Psittacosaurus lujiatunensis* (PKUP V1050), samples 121, 122 and 123 from *Caudipteryx zoui* (PMOL AD00020), samples 11, 13 and 14 from an undescribed enantiornithine (CUGB P1201), samples 30 and 33 from an undescribed enantiornithine (CUGB G20120001), and samples 64–69, 89 and 90 from an undescribed ornithurine bird (CUGB G20100053) were removed from the database. Samples are colour coded as in Fig. 4, main text and adjusted *n* for subsampling analysis follows: extant mammal hair (blue, $n = 719$), skin from extant (dark green, $n = 742$) and extinct (light green, $n = 605$) lepidosaurian, testudine and archosaurian species, feathers in basal Paraves (yellow, $n = 1,212$), *Confuciusornis* and crown-ward extinct avialan taxa (orange, $n = 1,376$), extant Aves (bright red, $n = 3,294$) and flightless palaeognath birds (dark red, $n = 107$). Colours of silhouettes correspond with colours in scatterplots. Black indicated unsampled taxa or integumentary type (e.g., bristle structures on the tail of *Psittacosaurus*).

LETTER

doi:10.1038/nature12961

The genomic landscape of Neanderthal ancestry in present-day humans

Sriram Sankararaman[1,2], Swapan Mallick[1,2], Michael Dannemann[3], Kay Prüfer[3], Janet Kelso[3], Svante Pääbo[3], Nick Patterson[1,2] & David Reich[1,2,4]

Genomic studies have shown that Neanderthals interbred with modern humans, and that non-Africans today are the products of this mixture[1,2]. The antiquity of Neanderthal gene flow into modern humans means that genomic regions that derive from Neanderthals in any one human today are usually less than a hundred kilobases in size. However, Neanderthal haplotypes are also distinctive enough that several studies have been able to detect Neanderthal ancestry at specific loci[1,3–8]. We systematically infer Neanderthal haplotypes in the genomes of 1,004 present-day humans[9]. Regions that harbour a high frequency of Neanderthal alleles are enriched for genes affecting keratin filaments, suggesting that Neanderthal alleles may have helped modern humans to adapt to non-African environments. We identify multiple Neanderthal-derived alleles that confer risk for disease, suggesting that Neanderthal alleles continue to shape human biology. An unexpected finding is that regions with reduced Neanderthal ancestry are enriched in genes, implying selection to remove genetic material derived from Neanderthals. Genes that are more highly expressed in testes than in any other tissue are especially reduced in Neanderthal ancestry, and there is an approximately fivefold reduction of Neanderthal ancestry on the X chromosome, which is known from studies of diverse species to be especially dense in male hybrid sterility genes[10–12]. These results suggest that part of the explanation for genomic regions of reduced Neanderthal ancestry is Neanderthal alleles that caused decreased fertility in males when moved to a modern human genetic background.

To search systematically for Neanderthal haplotypes, we developed a method based on a conditional random field[13] (CRF) that combines information from three features of genetic variation that are informative of Neanderthal ancestry (Supplementary Information section 1 and Extended Data Fig. 1). The first is the allelic pattern at a single nucleotide polymorphism (SNP): if a non-African individual carries a derived allele seen in Neanderthals but absent from the west-African Yoruba from Ibadan, Nigeria (YRI), the allele is likely to originate from Neanderthals. The second is high sequence divergence of the non-African haplotype to all YRI haplotypes but low divergence to Neanderthal. The third is a haplotype length consistent with interbreeding 37–86-thousand years ago[14]. We trained the CRF using simulations[15], and established its robustness to deviations from the assumed demography (Supplementary Information section 2).

We screened for Neanderthal haplotypes in the 1000 Genomes Project Phase 1 (1KG) data[9], using the Altai Neanderthal genome of 52-fold average coverage to determine alleles present in Neanderthals[2], a six-primate consensus to determine ancestral alleles[16], and 176 YRI genomes as a reference panel assumed to harbour no Neanderthal ancestry (Fig. 1a). Table 1 reports the mean and standard deviation across individuals of the fraction of their ancestry confidently inferred to be Neanderthal (probability > 90%). Figure 1b and Extended Data Fig. 2 plot the fraction of European ($n = 758$) and east-Asian ($n = 572$) haplotypes that descend from Neanderthals at each genomic location (Supplementary Information section 3). We created a tiling path of inferred Neanderthal haplotypes that spans 1.1 gigabases (Gb) over 4,437 contigs (Supplementary Information section 4), thus filling in gaps in the Neanderthal sequence over a number of repetitive regions that cannot be reconstructed from short ancient DNA fragments (Extended Data Fig. 3).

Four features of the Neanderthal introgression map suggest that it is producing reasonable results. First, when we infer Neanderthal ancestry using low-coverage data from Croatian Neanderthals[1] we obtain correlated inferences (Spearman rank correlation $\rho = 0.88$ in Europeans; Supplementary Information section 3). Second, in the African Luhya in Webuye, Kenya (LWK), the proportion of the genome inferred to be Neanderthal is 0.08%, an order of magnitude smaller than in non-African populations (Table 1). Third, the proportion of the genome with confidently inferred Neanderthal ancestry has a mean of 1.38% in east-Asian and 1.15% in European populations (Table 1), consistent with previous reports of more Neanderthal ancestry in east-Asian than

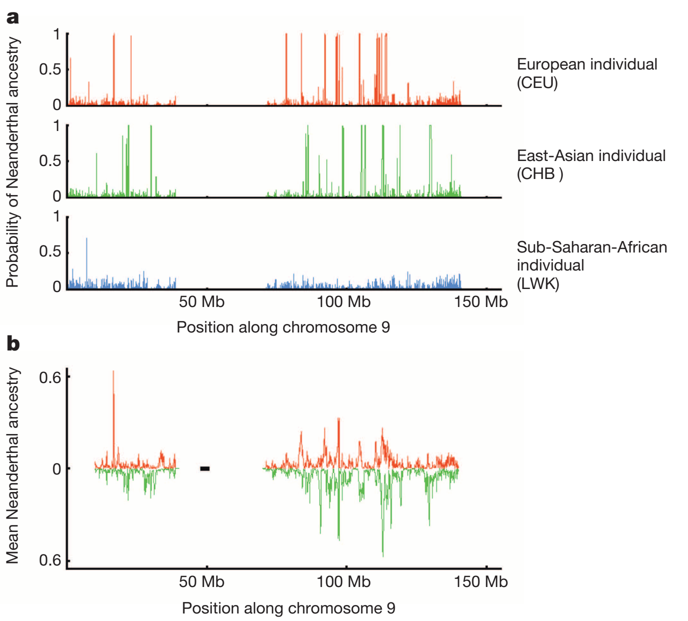


Figure 1 | Maps of Neanderthal ancestry. **a**, Individual maps; the marginal probability of Neanderthal ancestry for one European-American, one east-Asian and one sub-Saharan-African phased genome across chromosome 9. **b**, Population maps; estimate of the proportion of Neanderthal ancestry in European individuals (red) and east-Asian individuals (green), averaged across all individuals from each population in non-overlapping 100-kb windows on chromosome 9. The black bar denotes the coordinates of the centromere. The plot is limited to segments of the chromosome that pass filters (see Supplementary Information section 8). CEU, residents of Utah, US, with northern and western European ancestry (from the Centre d'Etude du Polymorphisme Humain (CEPH) collection); CHB, Han Chinese in Beijing, China; LWK, African Luhya in Webuye, Kenya.

[1]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. [3]Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany. [4]Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA.

Table 1 | Genome-wide estimates of Neanderthal ancestry

Region	Population	Number of individuals	Neanderthal ancestry on autosomes (%)	Neanderthal ancestry on the X chromosome (%)
Europe	CEU	85	1.17 ± 0.08	0.21 ± 0.17
	FIN	93	1.20 ± 0.07	0.19 ± 0.14
	GBR	89	1.15 ± 0.08	0.20 ± 0.15
	IBS	14	1.07 ± 0.06	0.23 ± 0.18
	TSI	98	1.11 ± 0.07	0.25 ± 0.20
East Asia	CHB	97	1.40 ± 0.08	0.30 ± 0.21
	CHS	100	1.37 ± 0.08	0.27 ± 0.21
	JPT	89	1.38 ± 0.10	0.26 ± 0.21
America	CLM	60	1.14 ± 0.12	0.22 ± 0.16
	MXL	66	1.22 ± 0.09	0.21 ± 0.15
	PUR	55	1.05 ± 0.12	0.20 ± 0.15
Africa	LWK	97	0.08 ± 0.02	0.04 ± 0.07
	ASW	61	0.34 ± 0.22	0.07 ± 0.11

For each computationally phased genome in each population, we estimated the probability of Neanderthal ancestry at each SNP and the fraction of autosomal and X chromosome SNPs that are confidently of Neanderthal origin in each individual (marginal probability >90%). The table reports the average and standard deviation of this statistic across individuals within each population. ASW, people with African ancestry in Southwest United States; CEU, Utah residents with northern and western European ancestry (from the Centre d'Etude du Polymorphisme Humain (CEPH) collection); CHB, Han Chinese in Beijing, China; CHS, Han Chinese in South China; CLM, Colombians in Medellin, Colombia; FIN, Finnish in Finland; GBR, British from England and Scotland, UK; IBS, Iberian populations in Spain; JPT, Japanese in Tokyo, Japan; LWK, African Luhya in Webuye, Kenya; MXL, people with Mexican ancestry in Los Angeles, California; PUR, Puerto Ricans in Puerto Rico; TSI, Toscani in Italy.

in European populations[7,17]. Fourth, the standard deviation in Neanderthal ancestry among individuals from within the same population is 0.06–0.10%, in line with theoretical expectation (Supplementary Information section 3), showing that Neanderthal ancestry calculators that estimate differences on the order of a per cent[18] are largely inferring statistical noise.

The Neanderthal introgression map reveals locations where Neanderthal ancestry is inferred to be as high as 62% in east-Asian and 64% in European populations (Fig. 1b and Extended Data Fig. 2). Several of these regions provide evidence of positive selection if we assume a model in which the distribution of Neanderthal ancestry has been governed by neutral drift; however, this assumption is problematic in light of the evidence for widespread negative selection against Neanderthal ancestry reported below (Supplementary Information section 5). To further explore whether Neanderthal alleles may have been affected by positive selection, we examined the 5% of genes with the highest inferred Neanderthal ancestry. We do not detect tissue-specific expression patterns; however genes involved in keratin filament formation and some other biological pathways are significantly enriched in Neanderthal ancestry in European populations, east-Asian populations, or both (Extended Data Table 1 and Supplementary Information section 6). Thus, Neanderthal alleles that affect skin and hair may have helped modern humans to adapt to non-African environments. We also investigated the relevance of Neanderthal alleles to present-day human biology by identifying alleles of Neanderthal origin (Supplementary Information section 7), and overlapping this list with alleles that have been associated with phenotypes of medical relevance[19,20]. We identify alleles of Neanderthal origin that affect lupus, biliary cirrhosis, Crohn's disease, optic-disk size, smoking behaviour, IL-18 levels and type 2 diabetes[20] (Extended Data Table 2).

The most striking feature of the introgression map is large 'deserts' of Neanderthal ancestry: on a 10-megabase (Mb) scale on the autosomes, there are 4 windows in European and 14 in east-Asian populations with Neanderthal ancestry $< 0.1\%$ (Extended Data Fig. 2 and Supplementary Information section 8). Two analyses show that these deserts are not artefacts of reduced power to detect ancestry. First, these regions were detected using a probability threshold for calling a segment as Neanderthal of $>25\%$, which results in a much higher sensitivity to true segments of Neanderthal ancestry than does a threshold of $>90\%$ (Supplementary Information section 3). Second, when we estimate Neanderthal ancestry in regions of low recombination rate where Neanderthal haplotypes are longer so that we have more power to detect them, we see a decreased Neanderthal ancestry proportion, opposite to the expectation from increased power ($\rho = 0.221$, $P = 4.4 \times 10^{-4}$ in European individuals; $\rho = 0.226$, $P = 1.9 \times 10^{-4}$ in east-Asian individuals) (Supplementary Information section 8). As we also observe multi-megabase regions of increased Neanderthal ancestry, part of the explanation for the ancestry deserts is likely to be small population sizes shortly after interbreeding (Supplementary Information section 8). However, selection also seems to have contributed to Neanderthal ancestry deserts, as we detect a correlation to functionally important regions (below).

To explore whether selection provides part of the explanation for regions of reduced Neanderthal ancestry, we tested for a correlation of Neanderthal ancestry to a previously published 'B statistic', in which low B implies a high density of functionally important elements[21]. We find that low B is significantly correlated to low Neanderthal ancestry: $\rho = 0.32$ in European populations ($P = 4.9 \times 10^{-87}$) and $\rho = 0.31$ in east-Asian populations ($P = 3.88 \times 10^{-68}$) (Fig. 2 and Supplementary Information section 8). This is not an artefact of reduced power, as there is expected to be reduced genetic variation in regions of low B which should make introgressed Neanderthal haplotypes stand out more clearly (Extended Data Table 3 and Supplementary Information section 2). We also estimated Neanderthal ancestry in quintiles of B statistic using an approach that is not biased by varying mutation rates, recombination rates, or genealogical tree depth[22], and confirmed that the quintile with the highest B has significantly higher Neanderthal

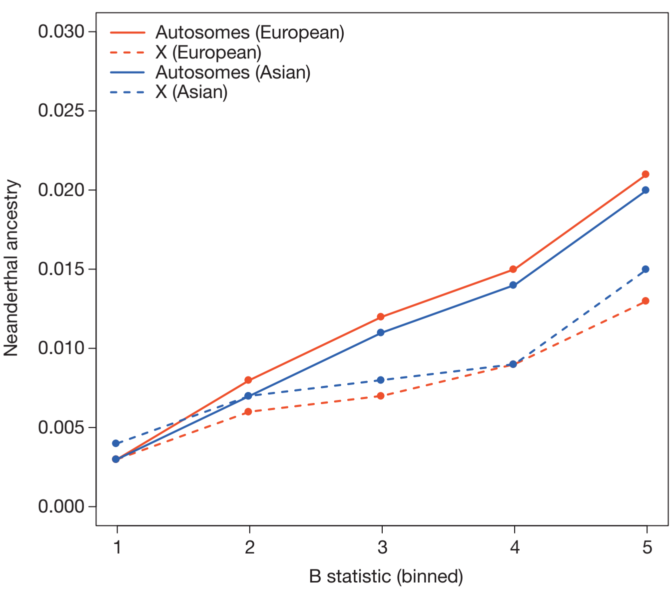


Figure 2 | Functionally important regions are deficient in Neanderthal ancestry. The median of the proportion of Neanderthal ancestry (estimated as the average over the marginal probability of Neanderthal ancestry assigned to each individual allele at a SNP) within quintiles of a B statistic that measures proximity to functionally important regions (1-low, 5-high). We show results on the autosomes and the X chromosome, and in European and east-Asian populations.

Table 2 | Enrichment of tissue-specific genes in regions deficient in Neanderthal ancestry

Tissue	European whole genome	European X chromosome	European autosomes	East-Asian whole genome	East-Asian X chromosome	East-Asian autosomes
Adipose	0.93	1	0.81	0.99	1	0.95
Adrenal	0.5	NA	0.5	0.42	NA	0.42
Blood	0.99	0.98	0.99	0.94	0.73	0.94
Brain	1	1	1	1	1	1
Breast	0.98	0.63	0.99	1	0.94	1
Colon	0.64	0.77	0.63	0.94	0.97	0.89
Heart	0.99	0.71	0.99	0.8	0.57	0.81
Kidney	1	0.15	1	1	0.08	1
Liver	0.99	0.99	0.99	1	0.86	1
Lung	0.96	0.64	0.96	0.99	0.87	0.99
Lymph	0.88	0.62	0.9	0.99	0.51	0.99
Ovary	0.84	0.95	0.81	0.62	0.91	0.58
Prostate	1	0.79	1	1	0.73	1
Muscle	0.95	0.7	0.95	0.83	0.1	0.88
Testes	0.0095	0.13	0.016	0.018	0.039	0.055
Thyroid	0.86	0.62	0.88	0.87	0.94	0.86

Tissue-specific genes (defined as those that are significantly more highly expressed in the specified tissue than in any of the 15 other tissues) are compared to all other expressed genes in that tissue. Of the 16 tissues tested, only testes-specific genes are significantly enriched in the regions deficient in Neanderthal ancestry, defined as locations in which all sites across all individuals are assigned a marginal probability of Neanderthal ancestry of <10% (47% of genes in European individuals and 52% of genes in east-Asian individuals fall into this category). NA, no tissue-specific genes for this tissue on the X chromosome. Units are *P* values.

ancestry than the other quintiles ($P = 7 \times 10^{-4}$) (Extended Data Table 4 and Supplementary Information section 9).

The largest deserts of Neanderthal ancestry are on the X chromosome, where the mean Neanderthal ancestry is about a fifth of the autosomes (Table 1). The power of our CRF to detect Neanderthal ancestry is higher on the X chromosome than on the autosomes (Extended Data Table 5 and Supplementary Information section 2), implying that this observation cannot be an artefact of reduced power. At least some of the reduction in Neanderthal ancestry that we observe on the X chromosome must be due to selection, since—just as on the autosomes—we observe that Neanderthal ancestry is positively correlated with B statistic ($\rho = 0.276, P = 3.1 \times 10^{-4}$ for European populations; $\rho = 0.176$, $P = 0.02$ for east-Asian populations) (Fig. 2 and Supplementary Information section 8). Studies in many species have shown that genes responsible for reduced male fertility disproportionally map to the X chromosome (refs 10–12). We reasoned that this 'large X effect'[23] could explain why the X chromosome was more resistant to introgression of Neanderthal ancestry than the autosomes.

If male hybrid sterility is contributing to our observations, a prediction is that the responsible genes will be disproportionally expressed in testes[24]. To test this hypothesis, we analysed gene transcripts from 16 human tissues[25] and defined 'tissue-specific' genes as those with a significantly higher expression level in that tissue than any other. We found that only genes that are specific to testes were enriched in regions of low Neanderthal ancestry (when compared with all other genes expressed in the same tissue). This effect remained significant after permuting gene annotations while preserving the correlation structure between Neanderthal ancestry and gene expression ($P = 0.0095$ in European populatons; $P = 0.018$ in east-Asian populations) (Table 2 and Supplementary Information section 6). However, hybrid sterility is not the only factor responsible for selection against Neanderthal material, as Neanderthal ancestry is also depleted in conserved pathways such as RNA processing ($P < 0.05$; Extended Data Table 2 and Supplementary Information section 6).

We have shown that interbreeding of Neanderthals and modern humans introduced alleles onto the modern human genetic background that were not tolerated, which probably resulted in part from their contributing to male hybrid sterility. The resulting reduction in Neanderthal ancestry was quantitatively large: in the fifth of the genome with highest B, Neanderthal ancestry is 1.54 ± 0.15 times the genome-wide average (Extended Data Table 4 and Supplementary Information section 9)[22]. If we assume that this subset of the genome was unaffected by selection, this implies that the proportion of Neanderthal ancestry shortly after introgression must have been >3% rather than the approximately 2% seen today. The large effect of negative selection on present-day levels of Neanderthal ancestry may explain why the proportion of Neanderthal ancestry is significantly higher in present-day east-Asian than in European populations (Table 1)[7,17]; there is evidence that east-Asian population sizes have been smaller than European populations for some of the time since their separation[26], which could have resulted in less efficient selection to remove Neanderthal-derived deleterious alleles. The evidence for male hybrid sterility is particularly remarkable when compared with mixed populations of present-day humans in which no convincing signals of selection against alleles inherited from one of the mixing populations have been found despite high power to detect such effects[27]. Thus, although the time of separation between Neanderthals and modern humans was only about five times larger than that between present-day European and west-African populations[2], the biological incompatibility was far greater. A potential explanation is the 'snowball effect', whereby hybrid sterility genes are expected to accumulate in proportion to the square of the substitutions between two taxa because two interacting loci need to change to produce an incompatibility ('Dobzhansky-Muller incompatibilities')[28]. An important direction for future work is to explore whether similar phenomena have affected other interbreeding events between diverged hominin groups.

METHODS SUMMARY

We infer the distribution of Neanderthal ancestry at each allele in a phased test genome using a conditional random field (CRF)[13]. The data we use consist of a panel of African reference genomes, the high-coverage Neanderthal genome, and a genetic map. For a set of CRF parameters, we can compute the probability of Neanderthal ancestry at each allele using the forward–backward algorithm (Supplementary Information section 1). Parameter estimation in the CRF needs training data (haplotypes labelled with true Neanderthal ancestries). We estimated the CRF parameters from data simulated under a demography relating European individuals, west-African individuals and Neanderthals (Supplementary Information section 2). We assessed the robustness of the CRF by measuring its false discovery rate in simulated demographic models that are different from the ones used in the simulation to train the CRF (Supplementary Information section 2). We also assessed the robustness to errors in phasing and in the genetic map (Supplementary Information section 2). For some analyses we estimate Neanderthal ancestry proportion based on all alleles where the marginal probability of Neanderthal ancestry as inferred by the CRF is greater than a specified threshold. For other analyses, we compute it as the average marginal probability of Neanderthal ancestry.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 September; accepted 18 December 2013.

Published online 29 January 2014.

1. Green, R. E. *et al.* A draft sequence of the Neanderthal genome. *Science* **328,** 710–722 (2010).
2. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505,** 43–49 (2014).

3. Abi-Rached, L. *et al.* The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* **334,** 89–94 (2011).
4. Mendez, F. L., Watkins, J. C. & Hammer, M. F. A haplotype at *STAT2* Introgressed from neanderthals and serves as a candidate of positive selection in Papua New Guinea. *Am. J. Hum. Genet.* **91,** 265–274 (2012).
5. Mendez, F. L., Watkins, J. C. & Hammer, M. F. Neanderthal origin of genetic variation at the cluster of OAS immunity genes. *Mol. Biol. Evol.* **30,** 798–801 (2013).
6. Yotova, V. *et al.* An X-linked haplotype of Neanderthal origin is present among all non-African populations. *Mol. Biol. Evol.* **28,** 1957–1962 (2011).
7. Wall, J. D. *et al.* Higher levels of neanderthal ancestry in East Asians than in Europeans. *Genetics* **194,** 199–209 (2013).
8. Lachance, J. *et al.* Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. *Cell* **150,** 457–469 (2012).
9. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
10. Tucker, P. K., Sage, R. D., Wilson, A. C. & Eichler, E. M. Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. *Evolution* **46,** 1146–1163 (1992).
11. Good, J. M., Dean, M. D. & Nachman, M. W. A complex genetic basis to X-linked hybrid male sterility between two species of house mice. *Genetics* **179,** 2213–2228 (2008).
12. Presgraves, D. C. Sex chromosomes and speciation in *Drosophila*. *Trends Genet.* **24,** 336–343 (2008).
13. Lafferty, J., McCallum, A. & Pereira, F. C. N. Conditional random fields: probabilistic models for segmenting and labeling sequence data. *Proc. 18th Int. Conf. Machine Learn.* 282–289 (2001).
14. Sankararaman, S., Patterson, N., Li, H., Paabo, S. & Reich, D. The date of interbreeding between Neanderthals and modern humans. *PLoS Genet.* **8,** e1002947 (2012).
15. Hellenthal, G. & Stephens, M. msHOT: modifying Hudson's ms simulator to incorporate crossover and gene conversion hotspots. *Bioinformatics* **23,** 520–521 (2007).
16. Paten, B., Herrero, J., Beal, K., Fitzgerald, S. & Birney, E. Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res.* **18,** 1814–1828 (2008).
17. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338,** 222–226 (2012).
18. Durand, E. Y. *Neanderthal Ancestry Estimator* White paper 23-05 http://23andme.https.internapcdn.net/res/pdf/hXitekfSJe1lcly7-Q72XA_23-05_Neanderthal_Ancestry.pdf (23andMe, 2011).
19. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106,** 9362–9367 (2009).
20. The SIGMA Type 2 Diabetes Consortium.. Sequence variants in *SLC16A11* are a common risk factor for type 2 diabetes in Mexico. *Nature* http://dx.doi.org/10.1038/nature12828 (25 December 2014).
21. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet.* **5,** e1000471 (2009).
22. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468,** 1053–1060 (2010).
23. Coyne, J. A. O. H. A. *Speciation and Its Consequences* (eds Otte, D. & Endler, J. A.) 180–207 (Sinauer Associates, 1989).
24. Wu, C.-I. & Davis, A. W. Evolution of postmating reproductive isolation: the composite nature of Haldane's rule and its genetic basis. *Am. Nat.* **142,** 187–212 (1993).
25. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22,** 1775–1789 (2012).
26. Keinan, A., Mullikin, J. C., Patterson, N. & Reich, D. Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nature Genet.* **39,** 1251–1255 (2007).
27. Bhatia, G. *et al.* Genome-wide scan of 29,141 African Americans finds no evidence of selection since admixture. Preprint at http://arxiv.org/pdf/1312.2675.pdf (2013).
28. Orr, H. A. & Turelli, M. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution* **55,** 1085–1094 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Briggs, P. Moorjani, M. Przeworski, D. Presgraves and A. Williams for critical comments, and K. Kavanagh for help with Extended Data Fig. 2. We are grateful for support from the Presidential Innovation Fund of the Max Planck Society, NSF HOMINID grant 1032255 and NIH grant GM100233. S.S. was supported by a post-doctoral fellowship from the Initiative for the Science of the Human Past at Harvard University. D.R. is a Howard Hughes Medical Institute Investigator.

Author Contributions S.S., N.P., S.P. and D.R. conceived of the study. S.S., S.M. M.D., K.P., J.K. and D.R. performed analyses. J.K., S.P., N.P. and D.R. supervised the study. S.S. and D.R. wrote the manuscript with help from all co-authors.

Author Information The tiling path of confidently inferred Neanderthal haplotypes, as well as the Neanderthal introgression map, can be found at http://genetics.med.harvard.edu/reichlab/Reich_Lab/Datasets.html. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.S. (sankararaman@genetics.med.harvard.edu) or D.R. (reich@genetics.med.harvard.edu).

METHODS

Conditional random field for inferring Neanderthal local ancestry. For a haploid genome in a test population that carries Neanderthal ancestry (for example, Europeans), given the allelic states of a sequence of SNPs along this haplotype, we wish to infer the ancestral state of the allele at each SNP; specifically, whether it has entered modern humans through Neanderthal gene flow. In addition to the test haplotype, the data analysed consist of a panel of haplotypes from the sub-Saharan-African Yoruba (YRI) who we assume harbour no Neanderthal ancestry[1]. To determine the allelic state of the Neanderthals, we used a high-coverage Neanderthal genome[2]. We determined the ancestral and derived allele at each SNP using a six-primate consensus sequence[16]. To estimate the genetic distance between adjacent SNPs, we used the Oxford combined linkage-disequilibrium map[29]. We specified the distribution of the unobserved Neanderthal ancestry states at each SNP given the observed genetic data as a conditional random field (CRF)[13]. We specified CRF feature functions that relate the observed data and the unobserved ancestral state at each SNP ('emission functions') as well as feature functions that relate the unobserved ancestral states at adjacent SNPs ('transition functions'). Thus, the model is a linear-chain CRF. The feature functions and their associated parameters fully specify the distribution of the unobserved ancestral states given the observed data. Given the parameters and the observed data, we were able to infer the marginal probability of Neanderthal ancestry at each SNP of the haploid genome. We computed the marginal probabilities efficiently using the forward–backward algorithm[13,30]. Supplementary Information section 1 presents the mathematical details.

Feature functions. The emission functions couple the unobserved ancestral state at a SNP to the observed features. We used two classes of emission functions.

The first class of emission functions captures information from the joint patterns observed at a single SNP across European and African individuals, and Neanderthals. These features are indicator functions that assume the value '1' when a specific pattern is observed at a SNP and '0' otherwise. We used feature functions that pick out two classes of allelic patterns. One of these features is 1 if at a given SNP, the test haplotype carries the derived allele, all the YRI haplotypes carry the ancestral allele, and either of the two Neanderthal alleles is derived. SNPs with this joint configuration have an increased likelihood of Neanderthal ancestry. In the CRF, an increased likelihood associated with this feature is reflected in the fact that the parameter is positive with a magnitude determined by the informativeness of the feature. The second feature is 1 if at a given SNP the test haplotype carries a derived allele that is polymorphic in the panel of African individuals but absent in the Neanderthal. SNPs with this joint configuration have a decreased likelihood of Neanderthal ancestry.

The second class of emission functions uses multiple SNPs to capture the signal of Neanderthal ancestry. Specifically, we compared the divergence of the test haplotype to the Neanderthal sequence to the minimum divergence of the test haplotype to all African haplotypes over non-overlapping 100-kilobase (kb) windows (the size scale we expect for Neanderthal haplotypes today based on the time of Neanderthal gene flow into modern humans[14]). In a region of the genome in which the test haplotype carries Neanderthal ancestry, we expect the test haplotype to be closer to the Neanderthal sequence than to most modern human sequences (albeit with a large variance), and we expect the pattern to be reversed outside these regions. We chose the derived allele at heterozygous sites so that this distance was effectively the minimum distance of the potentially introgressed test haplotype to one of the two Neanderthal haplotypes.

The transition feature function modulates the correlation of the ancestral states at adjacent SNPs. We defined this feature function as an approximation, at small genetic distance, to the log of the transition probabilities of a standard Markov process of admixture between two populations. This approximation makes parameter estimation in the CRF efficient.

Parameter estimation. To estimate the parameters of the CRF, we needed haplotypes labelled with Neanderthal ancestries; that is, training data. As we do not in fact know the true Neanderthal state in any individual, we estimated the CRF parameters on data simulated under a demographic model. We estimated parameters by maximizing the L2-regularized conditional log likelihood[30] using a limited-memory version of LBFGS (limited memory Broyden–Fletcher–Goldfarb–Shanno)[31]. We fixed the value of the parameter associated with the L2 penalty at 10 although a broad range of values seem to work in practice. We assumed a simple demographic model relating African and European individuals and Neanderthals with Neanderthal–modern human admixture occurring 1,900 generations ago[14] and a fraction of Neanderthal ancestry of 3%[1]. The model parameters were broadly constrained by the observed allele frequency differentiation between the west-African YRI and European-American CEU populations, and by the observed excess sharing of alleles between European and African populations relative to Neanderthals. The simulations incorporated hotspots of recombination[11] as well as the reduced power to detect low-frequency alleles from low-coverage sequencing data[12].

Validation of the conditional random field. We assessed the accuracy of the CRF to predict Neanderthal ancestry using simulated data. Given the marginal probabilities estimated by the CRF, we estimated the precision (fraction of predictions that are truly Neanderthal) and the recall (fraction of true Neanderthal alleles that are predicted) as we varied the threshold on the marginal probability for an allele to be declared Neanderthal. We also evaluated the accuracy when the haplotype phase needed to be inferred and when the genetic map had errors. As the CRF parameter estimation assumes a specific demographic model, we were concerned about the possibility that the inferences might be sensitive to the model assumed. We therefore varied each demographic parameter in turn and applied the CRF to data simulated under these perturbed models, fixing the parameters of the CRF to the estimates obtained under the original model. For each of these perturbed models, we evaluated the false discovery rate (defined as one minus the precision) when we restricted to sites at which the CRF assigns a marginal probability of at least 0.90. Supplementary Information section 2 presents the details.

Preparation of the 1000 Genomes data. We applied the CRF to the computationally phased haplotypes in each of the 13 populations in the 1000 Genomes Project[9] (1KG), excluding the west-African Yoruba (YRI). The CRF requires reference genomes from African individuals and Neanderthals. For the African population, we used 176 haplotypes from 88 YRI individuals. For the Neanderthal genome, for most analyses we used the genotypes called from the recently generated high-coverage Neanderthal sequence[2]. We restricted our analysis to sites passing the filters described in ref. 2 and for which the genotype quality score was ≥30. These filters discarded sites that are identified as repeats by the Tandem Repeat Finder annotation for hg19 from the University of California Santa Cruz (UCSC) genome browser (available at http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/simpleRepeat.txt.gz), that have Phred-scaled mapping quality scores of <30, or that map to regions in which the alignment is ambiguous or which fall within the upper or lower 2.5th percentile of the sample-specific coverage distribution (applied within the regions of unique mappability binned according to the GC content of the reference genome). For the mappability filter, we used the liberal filter that requires that at least 50% of all 35-mers that overlap a position do not map to any other position in the genome allowing up to one mismatch. We further restricted our analysis to sites that are biallelic across the Neanderthal and the 1000 Genomes Project samples. After filtering, we were able to analyse 26,493,206 SNPs on the autosomes and 817,447 SNPs on the X chromosome. For each haplotype analysed, we also restricted to the set of polymorphic sites in the population containing the haplotype. We obtained genetic distances from the Oxford combined linkage-disequilibrium map lifted over to hg19 coordinates[29]. For the X chromosome, we obtained an appropriate sex-averaged map by scaling the X chromosome linkage-disequilibrium-based map by two-thirds.

Statistics for measuring Neanderthal ancestry. We computed several statistics to summarize the Neanderthal ancestry inferred by the CRF. We estimated the proportion of an individual diploid genome that is confidently inferred to be Neanderthal as the fraction of sites for which the marginal probability is ≥90%. To assess variation in the proportion of Neanderthal ancestry along the genome, we computed the fraction of alleles across individuals with marginal probability greater than a specified threshold. This statistic is likely to be affected by variation in power along the genome. Hence, we also consider an estimate of the ancestry proportion obtained by averaging the marginal probability across individuals. Depending on the analyses, these statistics are estimated at a single SNP or in non-overlapping windows of a specified size.

To assess whether the predictions made by the CRF are sensible, we inferred Neanderthal ancestry using the low-coverage genome from the Vindija Neanderthals[1]. For this analysis, we restricted to sites at which there is at least one read with mapping quality score between 60 and 90 and base quality of at least 40. As a second validation analysis, we applied the CRF to the sub-Saharan-African Luhya (LWK), using the parameters estimated with non-African individuals. We empirically assessed the accuracy of the CRF on the 1000 Genomes Project data by assuming that LWK has no Neanderthal ancestry, that the false discovery rate in each non-African population is equal to the false discovery rate in LWK, and using the genome-wide proportion of Neanderthal ancestry estimated in ref. 2. (Supplementary Information section 3). We computed the theoretical standard deviation in the proportion of Neanderthal ancestry[32] assuming a pulse model of admixture with 2% Neanderthal ancestry followed by 2,000 generations of random mating, and 2.03 gigabases as the number of bases of the high-coverage Neanderthal genome that pass filters.

Tiling path of Neanderthal haplotypes. We identified Neanderthal haplotypes as runs of consecutive alleles along a test haplotype assigned a marginal probability of >90%. We filtered haplotypes smaller than 0.02 centiMorgans (cM). At each SNP that is covered by at least one such haplotype, we estimated the allelic state as the

consensus allele across the spanning haplotypes. See Supplementary Information section 4.

Functional analysis of introgressed alleles. We defined two subsets of consensus coding sequence (CCDS) genes[33]. We define a gene with 'low Neanderthal ancestry' as one in which all alleles across all individuals have a marginal probability ≤10%. We also require that the genes included in this analysis include at least ≥100 SNPs within a 100-kb window centred at its midpoint (this excluded genes at which the power to infer Neandertal ancestry is reduced due to a reduced number of informative sites). We define a gene with 'high Neanderthal ancestry' as one that is in the top 5% of CCDS genes ranked by the average marginal probability across individual haplotypes. See Supplementary Information section 6 for details.

Functional enrichment analysis. We tested for enrichment of Gene Ontology (GO)[34] categories in genes with low or high Neanderthal ancestry, using the hypergeometric test implemented in the FUNC package[35]. We report multiple-testing corrected *P* values estimated from 1,000 permutations for the GO enrichment analysis (family-wise error rate, FWER). Given the observed correlation between Neanderthal ancestry and B statistic[21], a concern is that the functional categories may not be randomly distributed with respect to B-statistic. To control for this, we assigned a B statistic to each gene (estimated as an average of the B statistic over the length of the gene) as well as a uniform random number. This resulted in 17,249 autosomal genes. Genes were binned into 20-equal sized bins based on the gene-specific B statistic. Within each bin, genes were sorted by Neanderthal ancestry and then by the random number. Genes ranked within the top 5% within each bin were used for the analysis. See Supplementary Information section 6 for details.

Identification of alleles born in Neanderthals and cross-correlation with association-study data. To infer whether an allele segregating in a present-day human population was introduced by Neanderthal gene flow, we defined a probable Neanderthal allele as one with marginal probability of ≥90% and a non-Neanderthal allele as having a marginal probability of ≤10%. A SNP at which all of the confident non-Neanderthal alleles as well as all alleles in YRI are ancestral and all of the confident Neanderthal alleles are derived is inferred to be of Neanderthal origin. This procedure allows for some false negatives in the prediction of the CRF. This procedure yields 97,365 Neanderthal-derived SNPs when applied to the predictions in European and east-Asian individuals. We downloaded the variants listed in the NHGRI (National Human Genome Research Institute) GWAS catalogue[19], retaining entries for which the reported association is a SNP with an assigned rs number, and for which the nominal *P* value is less than 5×10^{-8}. This resulted in 5,022 associations, which we then intersected with the Neanderthal-derived list. See Supplementary Information section 7 for details.

Identification of genomic regions deficient in Neanderthal ancestry. We measured the fraction of alleles across individuals and SNPs that are assigned a marginal probability greater than a chosen threshold measured within non-overlapping 10-megabase (Mb) windows. We chose a threshold of 25% as this threshold was found to lead to high recall in our empirical assessment (Supplementary Information section 3). We reported windows for which this statistic is <0.1%.

To understand the causes for variation in Neanderthal ancestry, we tested for correlation to a B statistic. Each SNP was annotated with the B statistic lifted over to hg19 coordinates. We assessed correlation between B and estimates of Neanderthal ancestry proportion at a nucleotide level as well as at different size scales, and separately on the autosomes and the X-chromosome (Supplementary Information section 8). We also used wavelet decomposition[36] to analyse the correlation of the inferred Neanderthal ancestry between European and east-Asian individuals at multiple size scales (Supplementary Information section 10). Figure 2 reports the relation between the mean marginal probability of Neanderthal ancestry across individuals and quintiles of B statistic at each SNP. To assess significance, we estimated Spearman's correlation ρ and standard errors on ρ using a block jackknife[37] with 10-Mb blocks (Supplementary Information section 8).

To understand the contribution of demography to variation in Neanderthal ancestry along the genome, we measured the coefficient of variation, at a 10-Mb scale, of the proportion of ancestry estimated as defined above. We then applied the CRF to data simulated under diverse demographic models and compared the coefficient of variation to the observed value (Supplementary Information section 8).

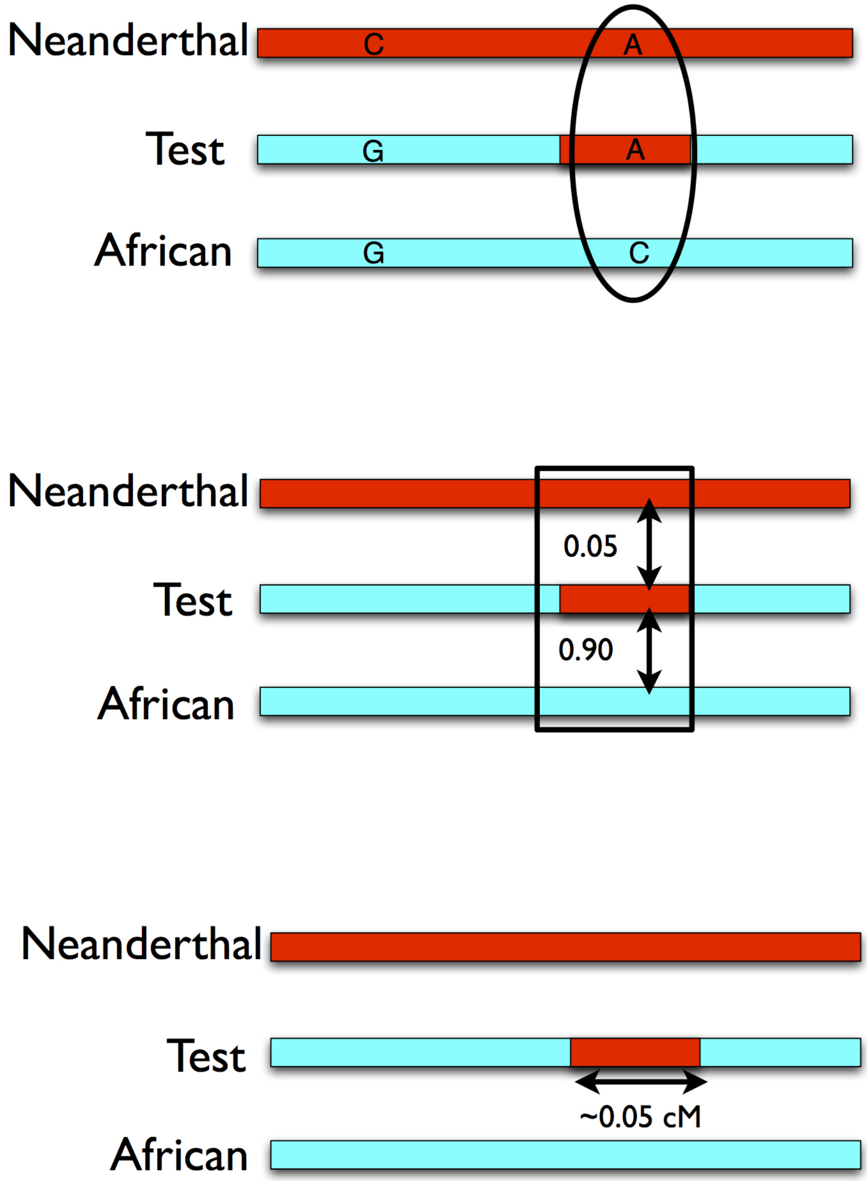
Assessment of the power to infer Neanderthal ancestry. To assess the power of the CRF to infer Neanderthal ancestry, we simulated data under diverse demographic models[38]. In one simulation series, we varied effective population size to approximate the effect of background selection and measured recall at a precision of 90%. In a second series, we assessed power on the X-chromosome versus the autosomes by matching the effective population size, recombination rate and mutation rate to estimated values for the X chromosome. See Supplementary Information section 2.

Unbiased estimate of the proportion of Neanderthal ancestry. To estimate the proportion of Neanderthal ancestry in an unbiased way, we divided the genome into quintiles of B, and estimated the proportion of Neanderthal ancestry using a statistic first published in ref. 22. This statistic measures how much closer Denisova is to a non-African individual than to an African individual, divided by the same quantity replacing the non-African individual with Neanderthal. We report the estimated proportion of Neanderthal ancestry in each quintile divided by the genome-wide mean and obtain standard errors using a block jackknife with 100 blocks.

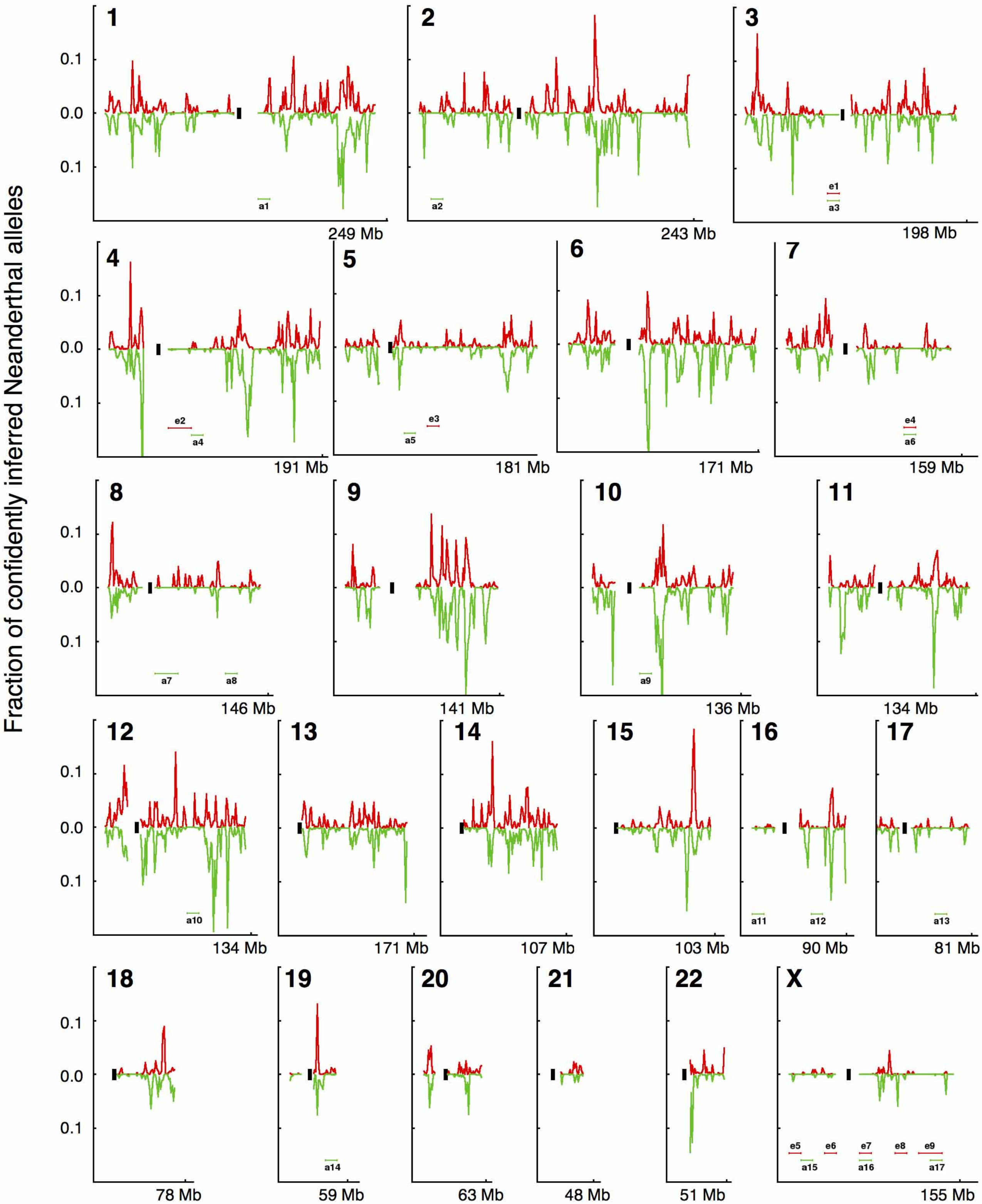
We analysed data from 27 deeply sequenced genomes: 25 present-day humans and the high-coverage Neanderthal and Denisova[17] genomes. For each, we required that sites passed the more stringent set of the two filters described in ref. 2, had a genotype quality of ≥45, and had an ancestral allele that could be determined based on comparison to chimpanzee and at least one of gorilla or orangutan. We computed a *Z*-score for the difference in the ancestry across the bin of highest B statistic versus the rest and used a Bonferroni correction for ten hypotheses (5 hypotheses based on which set of bins we merge and a two-sided test in each). In our main analysis, we analysed both transitions and transversions and pooled genomes for all non-African-individual samples. To establish robustness, we also analysed subsets of the data, repeating the analysis restricting to transversion SNPs, and then further restricting to transversion SNPs only in European and only in eastern-non-African individuals. See Supplementary Information section 9 for details.

Tissue-specific expression. We defined tissue-specific expression levels using the Illumina BodyMap 2.0 RNA-seq data[25], which contains expression data from 16 human tissues. We identified genes that are expressed in a tissue-specific manner using the DESeq package[39] and used a *P*-value cutoff of 0.05. We tested enrichment of tissue-specific genes in regions of high or low Neanderthal ancestry. A concern when testing for enrichment is that clustering of similarly expressed genes coupled with the large size of regions of low Neanderthal ancestry might lead to spurious signals of enrichment. Hence, we devised a permutation test that randomly rotates the annotations of genes (treating each chromosome as a circle) while maintaining the correlation within genes and within Neanderthal ancestry as well as between Neanderthal ancestry and genes. We tested enrichment on the whole genome, on the autosomes alone, and on the X chromosome alone. We generated 1,000 random rotations for each test except for the X chromosome for which we generated all possible rotations. We computed the fraction of permutations for which the *P* value of Fischer's exact test is at least as low as the observed *P* value (see Supplementary Information section 6).

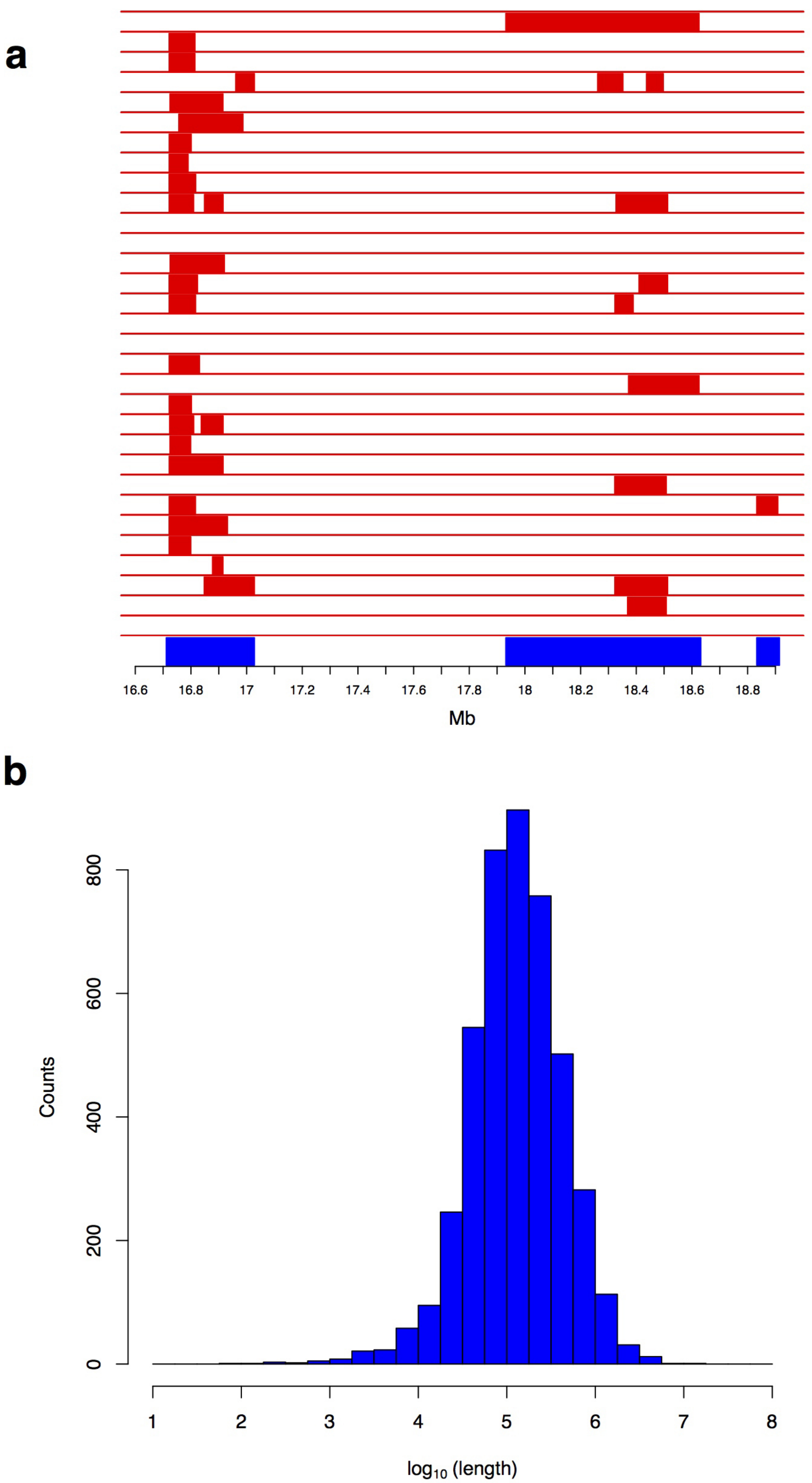
29. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310,** 321–324 (2005).
30. Sutton, C. & McCallum, A. in *Introduction to Statistical Relational Learning* (eds Getoor, L. & Taskar, B.) Ch. 4, 93–128 (MIT Press, 2007).
31. Byrd, R. H., Nocedal, J. & Schnabel, R. B. Representations of quasi-Newton matrices and their use in limited memory methods. *Mathematical Programming* **63,** 129–156 (1994).
32. Gravel, S. Population genetics models of local ancestry. *Genetics* **191,** 607–619 (2012).
33. Pruitt, K. D. *et al.* The consensus coding sequence (CCDS) project: identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res.* **19,** 1316–1323 (2009).
34. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature Genet.* **25,** 25–29 (2000).
35. Prüfer, K. *et al.* FUNC: a package for detecting significant associations between gene sets and ontological annotations. *BMC Bioinformatics* **8,** 41 (2007).
36. Percival, D. B. & Walden, A. T. *Wavelet Methods for Time Series Analysis*. (Cambridge Univ. Press, 2005).
37. Kunsch, H. R. The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17,** 1217–1241 (1989).
38. Hudson, R. R. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* **18,** 337–338 (2002).
39. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).



Extended Data Figure 1 | Three features used in the Conditional Random Field for predicting Neanderthal ancestry. Top (feature 1), patterns of variation at a single SNP. Sites at which a panel of sub-Saharan-African individuals carry the ancestral allele and in which the sequenced Neanderthal and the test haplotype carry the derived allele are likely to be derived from Neanderthal gene flow. Middle (feature 2), haplotype divergence patterns. Genomic segments in which the divergence of the test haplotype to the sequenced Neanderthal is low, whereas the divergence to a panel of sub-Saharan-African individuals is high, are likely to be introgressed. Bottom (feature 3), we searched for segments that have a length consistent with what is expected from Neanderthal-to-modern-human gene flow approximately 2,000 generations ago, corresponding to a size of about 0.05 cM = (100 cM per Morgan)/(2,000 generations).



Extended Data Figure 2 | Map of Neanderthal ancestry in 1000 Genomes European and east-Asian populations. For each chromosome, we plot the fraction of alleles confidently inferred to be of Neanderthal origin (probability >90%) in non-overlapping 1-Mb windows in Europeans (red) and in east Asians (green). Black bars denote the coordinates of the centromeres. We plot traces in non-overlapping 10-Mb windows that pass filters. We label 10-Mb-scale windows that are deficient in Neanderthal ancestry (e1–e9 (e, European), a1–a17 (a, Asian)) (see Supplementary Information section 8 for details).



Extended Data Figure 3 | Tiling path from confidently inferred Neanderthal haplotypes. **a**, Example tiling path at the *BNC2* locus on chromosome 9 in European individuals. Red, confidently inferred Neanderthal haplotypes in a subset of these individuals; blue, resulting tiling path. We identified Neanderthal haplotypes by scanning for runs of consecutive SNPs along a haplotype with a marginal probability >90% and requiring the haplotypes to be at least 0.02 cM long. **b**, Distribution of contig lengths obtained by constructing a tiling path across confidently inferred Neanderthal haplotypes. On merging Neanderthal haplotypes in each of the 1000 Genomes European and east-Asian populations, we reconstructed 4,437 Neanderthal contigs with median length 129 kb.

Extended Data Table 1 | Gene categories enriched or depleted in Neanderthal ancestry

Biological pathway (GO categorization)	Neandertal ancestry	Europe FWER	East Asian FWER
nucleic acid binding (molecular_function, GO:0003676)	Depleted	0.018	0.032
RNA processing (biological_process, GO:0006396)	Depleted	0.004	0.049
ribonucleoprotein complex (cellular_component, GO:0030529)	Depleted	<0.001	0.027
organelle part (cellular_component, GO:0044422)	Depleted	<0.001	0.037
intracellular organelle part (cellular_component, GO:0044446)	Depleted	<0.001	0.025
mRNA metabolic process (biological_process, GO:0016071)	Depleted	<0.001	0.014
nuclear lumen (cellular_component, GO:0031981)	Depleted	0.039	0.017
nuclear part (cellular_component, GO:0044428)	Depleted	0.005	0.022
keratin filament (cellular_component, GO:0045095)	Enriched	<0.001	<0.001

Enrichment of Gene Ontology categories in genes with depleted or elevated Neanderthal ancestry was assessed using the hypergeometric test implemented in the FUNC package. We report family-wise error rate (FWER) *P* values associated with each GO category (*P* values corrected for the testing of multiple categories).

Extended Data Table 2 | Neanderthal-derived alleles that have been associated with phenotypes in genome-wide association studies

rs id	Coordinates	Derived allele	Derived allele frequency (%)		Phenotype
			Europeans	East Asians	
rs12531711	7:128,617,466	G	10.03	0.17	Systemic lupus erythematosus, Primary biliary cirrhosis
rs3025343	9:136,478,355	A	8.44	0.00	Smoking behavior
rs7076156	10:64,415,184	A	26.52	8.74	Crohn's disease
rs12571093	10:70,019,371	A	16.35	14.86	Optic disc size
rs1834481	11:112,023,827	G	21.50	0.35	Interleukin-18 levels
rs11175593	12:40,601,940	T	1.98	3.32	Crohn's disease
rs75493593	17:6,945,087	T	1.85	12.06	
rs75418188	17:6,945,483	T	1.85	11.54	Type-2 Diabetes
rs117767867	17:6,946,330	T	1.85	11.54	

We identified alleles that are likely to have been introduced by Neanderthal gene flow (Supplementary Information section 7) and intersected these alleles with SNPs that have been shown to be associated with phenotypes (from the NHGRI GWAS catalogue[19] as well from a recent GWAS for type 2 diabetes[20]).

Extended Data Table 3 | Recall of the CRF as a function of the effective population size

Effective population size	Recall
2500	0.552 ± 0.009
5000	0.506 ± 0.009
7500	0.430 ± 0.006
10000	0.384 ± 0.006

Recall is computed at a precision of 90%. Standard errors of the recall are estimated by a block jackknife with 100 blocks.

Extended Data Table 4 | Unbiased estimate of the proportion of Neanderthal ancestry as a function of the B statistic

	11 Non-Africans (transitions + transversions)		**11 Non-Africans** (transversions only)		**4 Europeans** (transversions only)		**7 Eastern** (transversions only)	
	Est.	**Err.**	**Est.**	**Err.**	**Est.**	**Err.**	**Est.**	**Err.**
Quintile 1: B=0-0.63	0.641	0.304	0.672	0.316	0.472	0.397	0.778	0.317
Quintile 2: B=0.63-0.80	0.825	0.209	0.779	0.234	0.849	0.290	0.750	0.236
Quintile 3: B=0.80-0.88	0.578	0.248	0.745	0.298	0.987	0.349	0.647	0.297
Quintile 4: B=0.88-0.94	0.684	0.184	0.676	0.208	0.446	0.256	0.771	0.221
Quintile 5: B=0.94-1.00	1.537	0.152	1.445	0.164	1.502	0.185	1.419	0.177
B≥0.94 vs. B ≤ 0.94	Z=3.82		Z=3.02		Z=3.12		Z=2.58	
Correct for 10 hypotheses	P=0.00066		P=0.013		P=0.0090		P=0.049	

Estimates of the proportion of Neanderthal ancestry in quintiles of B statistics divided by the genome-wide proportion. We find a significant excess of Neanderthal ancestry in the quintile with the highest B statistic relative to the remaining four quintiles (significant after correcting for ten hypotheses).

Extended Data Table 5 | Recall of the CRF on the X chromosome versus the autosomes

	Recall
Autosomes	0.384 ± 0.006
X	0.495 ± 0.009

Simulations are carried out using parameters tailored to be appropriate to each of these compartments of the genome. Recall is computed at a precision of 90%. Standard errors of the recall are estimated by a block jackknife with 100 blocks.

doi:10.1038/nature12989

A dedicated circuit links direction-selective retinal ganglion cells to the primary visual cortex

Alberto Cruz-Martín[1,2], Rana N. El-Danaf[1,2], Fumitaka Osakada[3], Balaji Sriram[2], Onkar S. Dhande[1,2], Phong L. Nguyen[1,2], Edward M. Callaway[3], Anirvan Ghosh[4] & Andrew D. Huberman[1,2,3,5]

How specific features in the environment are represented within the brain is an important unanswered question in neuroscience. A subset of retinal neurons, called direction-selective ganglion cells (DSGCs), are specialized for detecting motion along specific axes of the visual field[1]. Despite extensive study of the retinal circuitry that endows DSGCs with their unique tuning properties[2,3], their downstream circuitry in the brain and thus their contribution to visual processing has remained unclear. In mice, several different types of DSGCs connect to the dorsal lateral geniculate nucleus (dLGN)[4–6], the visual thalamic structure that harbours cortical relay neurons. Whether direction-selective information computed at the level of the retina is routed to cortical circuits and integrated with other visual channels, however, is unknown. Here we show that there is a di-synaptic circuit linking DSGCs with the superficial layers of the primary visual cortex (V1) by using viral trans-synaptic circuit mapping[7,8] and functional imaging of visually driven calcium signals in thalamocortical axons. This circuit pools information from several types of DSGCs, converges in a specialized subdivision of the dLGN, and delivers direction-tuned and orientation-tuned signals to superficial V1. Notably, this circuit is anatomically segregated from the retino-geniculo-cortical pathway carrying non-direction-tuned visual information to deeper layers of V1, such as layer 4. Thus, the mouse harbours several functionally specialized, parallel retino-geniculo-cortical pathways, one of which originates with retinal DSGCs and delivers direction- and orientation-tuned information specifically to the superficial layers of the primary visual cortex. These data provide evidence that direction and orientation selectivity of some V1 neurons may be influenced by the activation of DSGCs.

Visual perception involves the activity of neurons in the cerebral cortex. The most direct route for visual information to reach the cortex is via the 'retino-geniculo-cortical pathway' consisting of retinal ganglion cells (RGCs), relay cells in the dLGN and neurons in primary visual cortex (V1) (Extended Data Fig. 1a–c)[9]. Recently, we and others discovered that direction-selective retinal ganglion cells (DSGCs) project to the dLGN and therein target a specific layer in the lateral 'shell'[4–6,10] (Fig. 1a–c). Hereafter we also refer to this layer as the DSGC-recipient zone or 'DSGC-RZ' (Fig. 1d).

Previous work showed that the dLGN shell receives input from the superior colliculus[11] and thus, like other thalamic compartments[12], neurons in the shell/DSGC-RZ may restrict their connections to subcortical networks, rather than participating in the retino-geniculo-cortical pathway. We infected neurons in the dLGN shell and a small portion of the dLGN core by injections of adeno-associated virus (AAV)-tdTomato (Fig. 1e). Within V1, tdTomato$^+$ axons were observed in deeper layers 4 and 6 and superficial layers 1 and 2 (Fig. 1f–i). Thus, neurons in the shell/DSGC-RZ probably include thalamocortical relay neurons, but it was unclear whether they target specific V1 layers.

To determine whether there is laminar specificity of mouse geniculocortical connections we injected retrograde tracers into different V1 layers (Fig. 2a–c) and analysed the position of the retrogradely labelled neurons in the dLGN (Fig. 2d–i) (Extended Data Fig. 2a–d). Injections of all V1 layers retrogradely labelled cells across the full width of the dLGN (Fig. 2a, d, g). By contrast, injections directed to V1 layer 4 preferentially labelled neurons in the dLGN core (Fig. 2b, e, h) and injections into superficial V1 layers 1 and 2 preferentially labelled neurons in the dLGN shell/DSGC-RZ (Fig. 2c, f, i) (Extended Data Fig. 3a–c). These laminar-specific patterns of retrograde labelling were independent of retinotopy or eye-specific connectivity (Extended Data Fig. 4) and together they indicate that cells in the dLGN core project to deeper V1, whereas cells in the dLGN shell/DSGC-RZ preferentially target superficial V1 (Extended Data Fig. 3d).

RGC axons form synapses onto the somas and dendrites of dLGN neurons[13], the latter of which are not entirely labelled using traditional retrograde tracing methods. Thus, we extended our exploration of the connections between DSGCs and thalamocortical relay neurons by using a glycoprotein-deleted rabies virus expressing mCherry (ΔG-RABV-mCherry) that infects neurons at the level of their axon terminals, leading to Golgi-like expression of mCherry throughout the infected cell. We note, however, that by itself ΔG-RABV-mCherry does not pass trans-synaptically[7,8]. Following injections of ΔG-RABV-mCherry into superficial V1 (Fig. 3a, b) we observed relay neurons in the dLGN shell expressing mCherry throughout their somas and dendritic arborizations (Fig. 3c, d). By performing these experiments in mice with genetically tagged DSGCs, we determined that the majority

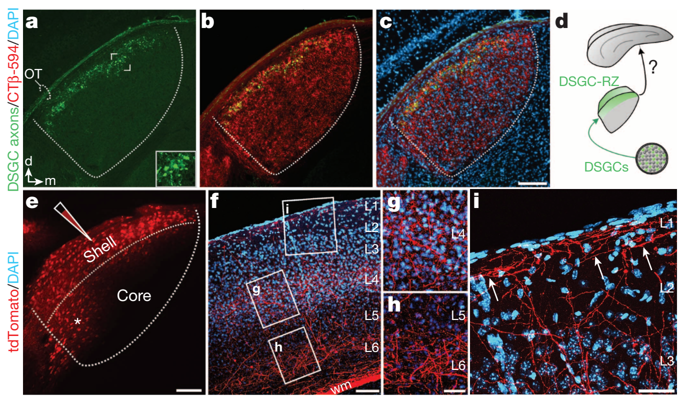


Figure 1 | The layer of the dLGN that receives input from DSGCs projects to V1. a, GFP$^+$ DSGC axons in dLGN. OT, optic tract. Inset, GFP$^+$ boutons. d, dorsal; m, medial. **b**, Merged GFP/CTβ−594 (all RGC axons) and **c**, with DAPI (4′,6-diamidino-2-phenylindole; cell nuclei). **a–c**, Scale bar, 125 µm. **d**, Summary. **e**, AAV2-tdTomato injection to dLGN. Asterisk, tdTomato$^+$ neurons in core. Scale bar, 100 µm. **f**, tdTomato$^+$ thalamocortical axons in V1. L1–L6, layers 1–6; wm, white matter. Scale bar, 150 µm. **g–i**, Higher magnification of L4 (**g**), L5 and L6 (**h**) and L1 and L2 (**i**). Arrows, axons in L1. Scale bars, **g**, **h**, 50 µm; **i**, 100 µm. Sixteen mice were used for these experiments.

[1]Department of Neurosciences, University of California, San Diego, California 92093, USA. [2]Neurobiology Section in the Division of Biological Sciences, University of California, San Diego, California 92093, USA. [3]Salk Institute for Biological Studies, La Jolla, California 92097, USA. [4]Neuroscience Discovery, F. Hoffman La Roche, 4070 Basel, Switzerland. [5]Department of Ophthalmology, University of California, San Diego, California 92093, USA.

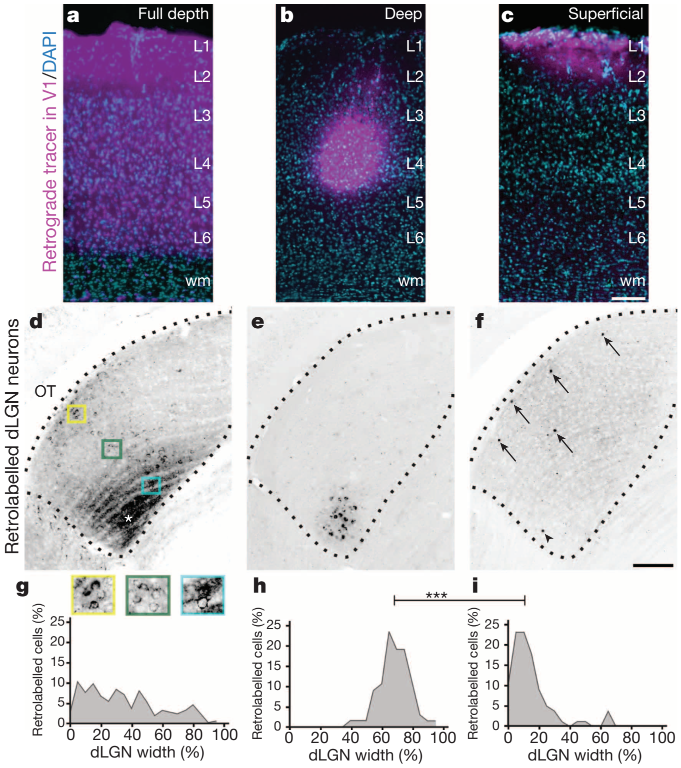


Figure 2 | Parallel, layer-specific thalamocortical circuits in the mouse. **a**–**c**, Tracer injections to all V1 layers (**a**), layer 4 (**b**) and superficial V1 (**c**). **d**–**f**, dLGN neurons labelled after full depth injection (**d**; asterisk, axons of L6 neurons; yellow box, neurons in shell; green box, middle; blue box, core) deep V1 injection (**e**) and superficial V1 injection (**f**; arrows, labelled cells; arrowhead, labelled cell outside shell). **g**–**i**, Position of retrogradely labelled cells along dLGN width. **g**, Full depth, 35.54 ± 1.62% (4 mice; $n = 232$ cells); **h**, Deep, 68.53 ± 1.22% (4 mice; $n = 69$ cells); **i**, Superficial, 14.68 ± 1.57% (4 mice; $n = 78$ cells). Superficial versus deep = $***P < 0.0001$; Tukey's multiple comparisons. **c**, Scale bar, 200 μm. **f**, Scale bar, 100 μm.

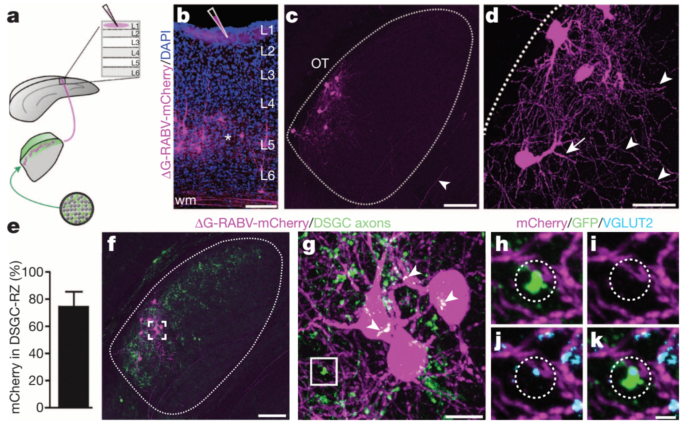


Figure 3 | DSGC axons contact thalamic relay neurons projecting to superficial V1. **a**, ΔG-RABV-mCherry injection to superficial V1, to infect axons of DSGC-RZ neurons. **b**, V1 injection. Asterisk, infected L5/6 neurons. Scale bar, 200 μm. **c**, **d**, mCherry$^+$ dLGN neurons. **c**, Arrowhead, axon. Scale bars, 75 μm. **d**, Arrow, proximal dendrites; arrowheads, distal dendrites. Scale bars, 25 μm. **e**, Percentage of mCherry$^+$ somas within GFP$^+$ DSGC-RZ (74.77 ± 11.62%; 8 mice, $n = 83$ cells) (Extended Data Fig. 4). **f**–**h**, DSGC axons and dLGN somas and dendrites. **g**, Magnified view of frame in **f**. Arrowheads, putative contact sites[13]. Scale bars; **f**, 125 μm; **g**, 20 μm. **h**–**k**, GFP, mCherry, VGLUT2 (per cent mCherry signal contacted by GFP$^+$/VGLUT2$^+$ profiles = 5.23 ± 1.39%; 4 mice; $n = 4$ cells). **k**, Scale bar, 2 μm.

of the mCherry$^+$ cells resided in the DSGC-RZ (Fig. 3e, f) (Extended Data Fig. 4) (Methods). We observed putative sites of contact between the axon terminals of DSGCs and the dendrites of mCherry$^+$ dLGN neurons (Fig. 3g), some of which contained VGLUT2 (Fig. 3h–k) (Extended Data Fig. 5), a presynaptic glutamate transporter that, in the dLGN, arises solely from RGC axon terminals[14]. Together, these experiments indicate that the neurons in the dLGN shell/DSGC-RZ that project to superficial V1 are contacted by, and probably receive synaptic input from, On-Off DSGCs.

To determine whether there is a bona fide di-synaptic circuit linking DSGCs to superficial V1, we used rabies-virus trans-synaptic network tracing[7,8]. We injected ΔG-RABV-mCherry into superficial V1 to infect neurons in the dLGN shell/DSGC-RZ via their presynaptic terminals and, in the same mice, we infected dLGN neurons with an AAV expressing the rabies glycoprotein and histone-tagged green fluorescent protein (AAV2-Glyco-hGFP) (Fig. 4a). In this experimental configuration, the double-infected RABV-mCherry$^+$/Glyco-GFP$^+$ dLGN relay neurons produce infectious ΔG-RABV-mCherry that propagates trans-synaptically to infect and label the RGCs that form synapses with them (Fig. 4a) (Methods). A small number of double-infected cells were present in the dLGN and these were always located in the shell/DSGC-RZ (Fig. 4b–e) (8 mice, $n = 21$ cells). Moreover, we performed these experiments in mice where GFP is selectively expressed by posterior-tuned On-Off DSGCs[5,6] (Extended Data Fig. 6) and we used immunohistochemical markers that recognize multiple On-Off DSGC types[15,16] to determine (1) whether DSGCs provide di-synaptic input to V1, and (2) if so, whether multiple DSGC types feed this pathway.

In mice with one to three double-infected dLGN neurons we observed 1-3 mCherry$^+$ RGCs, which is consistent with the convergence of mouse retinogeniculate connections[17]. An example of a trans-synaptically labelled RGC is shown in Fig. 4f, g. The cell has thin dendrites and looping arborizations and is bistratified in the On and Off sublayers of the inner plexiform layer (IPL) (Fig. 4g, and inset), features characteristic of On-Off DSGCs[2,3,5,6]. Also, the RGC is GFP$^+$ in a transgenic mouse where GFP is selectively expressed in posterior-tuned On-Off DSGCs[6] (Fig. 4h–j) (Extended Data Fig. 7). Another mCherry$^+$ DSGC trans-synaptically labelled from superficial V1 is displayed in Fig. 4k–m. The cell expresses cocaine- and amphetamine-regulated transcript (Cart), a marker of On-Off DSGCs[15,16]. Interestingly, trans-synaptic tracing from superficial V1 also labelled J-RGCs, which are asymmetric Off-type DSGCs tuned for upward motion[4] (Fig. 4n, o).

Of all the RGCs trans-synaptically labelled from superficial V1 we verified that the vast majority were DSGCs ($n = 27/28$ RGCs; 8 mice). 17/28 were genetically identified as posterior-tuned On-Off DSGCs[5,6] and 9/28 were Cart$^+$ but not GFP$^+$ and thus are On-Off DSGCs which are probably tuned to other cardinal axes of motion[15,16]. 1/28 was an upward tuned J-RGC[4], and one could not be classified. Thus, although we cannot conclude that only DSGCs contribute to this pathway, we find that several types of On-Off DSGCs as well as Off-type DSGCs provide di-synaptic input to superficial V1.

Do DSGCs also feed the classic retino-geniculo-cortical pathway into layer 4? We addressed this by injecting mice with ΔG-RABV-mCherry into V1 layer 4 and AAV2-Glyco-hGFP into the dLGN, then examining the RGC types trans-synaptically labelled with mCherry (Fig. 4p). In this regime, the mCherry$^+$/Glyco-hGFP$^+$ double-infected neurons resided in the dLGN core (Fig. 4q–t) (Extended Data Fig. 8) (7 mice, $n = 53$ cells) (Fig. 4t versus Fig. 4e; $***P < 0.0001$; two-tailed t-test). All the RGCs trans-synaptically labelled from layer 4 had large somas and broad, smooth monostratified dendrites, features characteristic of mouse alpha RGCs[18,19] (Fig. 4u, v) and/or they expressed SMI-32, a marker of alpha RGCs[18] (Fig. 4w–y) (7 mice, $n = 38$ RGCs) that is not expressed by On-Off DSGCs (Fig. 4z–bb). Moreover, none of the RGCs trans-synaptically infected from layer 4 exhibited DSGC morphologies, stratification patterns or molecular/genetic markers[4–6,15,16] (Fig. 4cc–ee) (0/38 RGCs examined; 7 mice). Combined with the results in Figs 2, 3, 4a–o, these data indicate that alpha, non-direction-tuned RGCs are

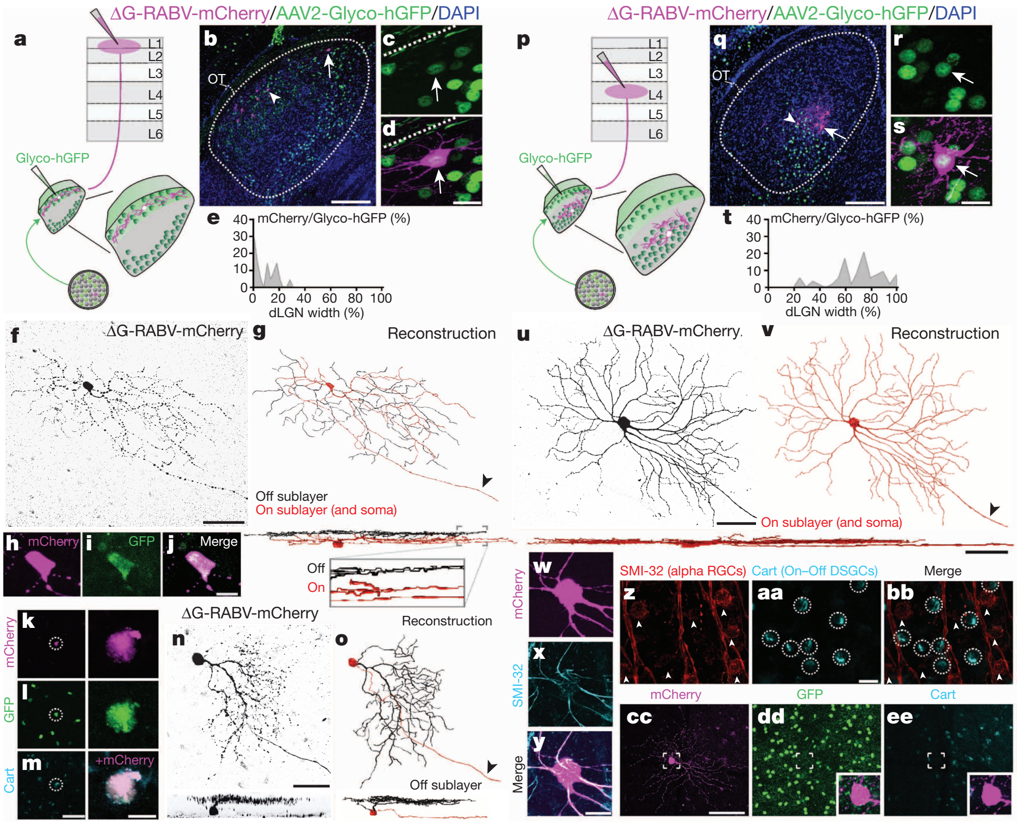


Figure 4 | Synaptic circuit linking DSGCs to superficial V1, and non-DSGCs to L4. **a**, Trans-synaptic tracing. **b**, Infected dLGN neurons. Arrow and arrowhead: double-infected cells; arrow is same cell as in **c**, **d**. Scale bar, 100 μm. **c**, **d**, Cell from **b**. Scale bar, 15 μm. Dashed line, lateral border. **e**, Distribution of double-infected dLGN cells = 9.29 ± 1.82% (8 mice, $n = 21$ cells). **f**, On-Off DSGC trans-synaptically labelled from superficial V1. **g**, On (red) and Off (black) dendrites. Arrowhead, axon. Scale bar, 50 μm. **h–j**, Cell (**f**) is GFP$^+$ On-Off DSGC[6]. Scale bar, 10 μm. **k–m**, Trans-synaptically labelled GFP$^+$ and Cart$^+$ DSGC[5] shown at low (left) and high (right) magnifications. **m**, Scale bar left, 75 μm; right, 10 μm. **n**, Trans-synaptically labelled J-RGC[4]. **o**, Off dendrites (black). Scale, 50 μm. **p**, Same as **a**, but layer 4 injection. **q–s**, Infected neurons in core; **q**, arrow, arrowhead: double-infected cells. Scale, 100 μm. **r**, **s**, Cell from **q** (arrow). **s**, Scale bar, 15 μm. **t**, Distribution of double-infected dLGN cells = 70 ± 2.65% (7 mice, $n = 53$ cells) ($P < 0.0001$ versus **e**; two-tailed *t*-test). **u**, **v**, Alpha RGC labelled from V1 layer 4. **u**, Scale bar, 100 μm; sideview, 50 μm. **w–y**, mCherry$^+$ RGC same as from **u**, **v**, is SMI-32$^+$. Scale bar, 20 μm. **z–bb**, SMI-32 and Cart. Scale bar, 25 μm. **cc–ee**, ΔG-RABV-mCherry$^+$ alpha RGC; lacks GFP[6] and Cart. Scale bar, 150 μm.

poised to influence layer 4 of V1 via neurons in the dLGN core, whereas DSGCs are poised to influence superficial V1 via neurons in the dLGN shell.

Next we asked what qualities of visual information are delivered by the dLGN to superficial V1. We forced a subset of dLGN neurons to express the calcium indicator GCaMP6[20] by injections of AAV2-Syn-GCaMP6 into the shell/DSGC-RZ (Fig. 5a–c) (5 mice). We then imaged the visually evoked calcium dynamics (Δ*F*/*F*) in thalamocortical axons that target superficial layers of V1 (Fig. 5d) using *in vivo* time-lapse two-photon microscopy[20] (Fig. 5e), while presenting the mice with drifting gratings of different orientations and directions (Fig. 5e–g) (Methods).

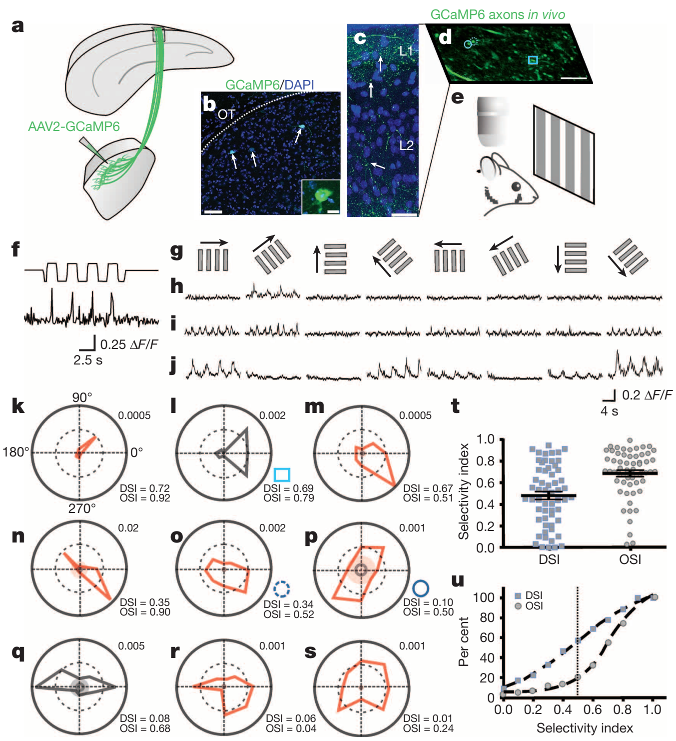


Figure 5 | *In vivo* imaging of visually evoked Ca^{2+} signals in thalamocortical axons. **a**, AAV2-GCaMP6 injection to dLGN shell. **b**, GCaMP6$^+$ neurons (arrows). Scale bar, 50 μm; inset, 10 μm. **c**, GCaMP6$^+$ dLGN axons, superficial V1. Arrows: varicosities. Scale, 50 μm. **d**, GCaMP6$^+$ axons, superficial V1. Circles, square in **d** correspond to polar plots **l**, **o**, **p**. Scale bar (**d**), 5 μm. **e**, *In vivo* imaging/visual stimulation. **f**, Visually evoked Ca^{2+} signal in thalamocortical axon (top trace: photodiode signal; bottom trace: Δ*F*/*F*). **g**, Directional stimuli (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°). **h–j**, Direction- (**h**, **i**) and orientation-tuned (**j**) varicosities. 5–8 trial average. **k–s**, Polar plots of F1 (red) or F2 (black) magnitude responses (Methods). Inner solid ring, average response to mean grey stimulus. Shaded, 3 standard deviations greater than the mean response to grey stimuli. Lower right of each plot, OSI/DSI. Upper right, Fourier amplitudes. **t**, DSI/OSI, all varicosities (5 mice, $n = 58$ varicosities). Mean ± s.e.m. **u**, Cumulative distributions: OSI (circles), DSI (squares).

Figure 5f shows an example of visually evoked calcium transients in a thalamocortical axon. Figure 5h, i shows two clear and marked examples of direction-tuned signals in dLGN axons located within superficial V1. The first example (Fig. 5h), responds only to gratings drifting at 45 degrees (polar plot Fig. 5k). The second example (Fig. 5i) is also direction-selective but is more broadly tuned (polar plot Fig. 5l). Interestingly, we also observed thalamocortical axons that were strongly tuned not for direction, but instead for orientation (Fig. 5j and Fig. 5n–q). Overall, a variety of strengths of direction and orientation tuning were observed, including some thalamocortical axons that were not tuned for either feature (Fig. 5r, s). Approximately 60% of all thalamocortical axons imaged had direction selectivity indices (DSIs) ≥ 0.4, a commonly used threshold for categorizing mouse DSGCs[21] and approximately $\sim$85% had orientation selectivity indices (OSIs) of ≥ 0.5 (Fig. 5t, u) (5 mice, $n = 58$ varicosities). Thus, the majority of visual information delivered to superficial V1 by neurons in the dLGN shell/DSGC-RZ is direction- and/or orientation-tuned.

Our results are the first to define a circuit relationship between DSGCs and visual cortex and they suggest that, in the mouse, direction-tuned and orientation-tuned V1 neurons may inherit their characteristic receptive field properties from DSGCs. Notably, such influence may be specific to neurons receiving thalamic excitation within superficial V1. Indeed, our results of trans-synaptic labelling of RGCs from V1 layer 4, combined with recent studies that recorded thalamic excitation in layer 4 (refs 22, 23), suggest that orientation and direction selectivity of neurons in deeper V1 arises from the convergence of retinal and thalamic afferents with centre-surround receptive fields, as opposed to direction- or orientation-tuned receptive fields.

Interestingly, we discovered that several different types of On-Off DSGCs as well as upward-tuned Off-DSGCs feed the retino-geniculo-superficial V1 circuit. We also found that some thalamocortical axons are tuned for specific axes of motion, whereas others are orientation-tuned[24,25]. Thus, the tuning of V1 neurons that receive input from neurons in the shell/DSGC-RZ may reflect the integration of individual or multiple DSGC types, as well as intra-cortical circuitry[26–28]. Together, our findings suggest that directional motion encoded at the level of the retina may influence the direction and orientation selectivity of some V1 neurons.

METHODS SUMMARY

All experimental protocols were conducted according to the National Institutes of Health (NIH) guidelines for animal research and approved by the Institutional Animal Care and Use Committee (IACUC) at the University of California, San Diego, USA.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 April; accepted 31 December 2013.

Published online 26 February; corrected online 19 March 2014 (see full-text HTML version for details).

1. Barlow, H. B. & Hill, R. Selective sensitivity to direction of movement in ganglion cells of the rabbit retina. *Science* **139,** 412–414 (1963).
2. Briggman, K. L., Helmstaedter, M. & Denk, W. Wiring specificity in the direction-selectivity circuit of the retina. *Nature* **471,** 183–188 (2011).
3. Wei, W. & Feller, M. B. Organization and development of direction-selective circuits in the retina. *Trends Neurosci.* **34,** 638–645 (2011).
4. Kim, I. J., Zhang, Y., Yamagata, M., Meister, M. & Sanes, J. R. Molecular identification of a retinal cell type that responds to upward motion. *Nature* **452,** 478–482 (2008).
5. Huberman, A. D. *et al.* Genetic identification of an On-Off direction-selective retinal ganglion cell subtype reveals a layer-specific subcortical map of posterior motion. *Neuron* **62,** 327–334 (2009).
6. Rivlin-Etzion, M. *et al.* Transgenic mice reveal unexpected diversity of on-off direction-selective retinal ganglion cell subtypes and brain structures involved in motion processing. *J. Neurosci.* **31,** 8760–8769 (2011).
7. Wickersham, I. R. *et al.* Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* **53,** 639–647 (2007b).
8. Osakada, F. *et al.* New rabies virus variants for monitoring and manipulating activity and gene expression in defined neural circuits. *Neuron* **71,** 617–631 (2011).
9. Chalupa, L. M. & Werner, J. S. *The Visual Neurosciences* (MIT Press, 2003).
10. Krahe, T. E., El-Danaf, R. N., Dilger, E. K., Henderson, S. C. & Guido, W. Morphologically distinct classes of relay cells exhibit regional preferences in the dorsal lateral geniculate nucleus of the mouse. *J. Neurosci.* **31,** 17437–17448 (2011).
11. Grubb, M. S. & Thompson, I. D. Biochemical and anatomical subdivision of the dorsal lateral geniculate nucleus in normal mice and in mice lacking the β2 subunit of the nicotinic acetylcholine receptor. *Vision Res.* **44,** 3365–3376 (2004).
12. Morin, L. P. & Blanchard, J. H. Forebrain connections of the hamster intergeniculate leaflet: comparison with those of ventral lateral geniculate nucleus and retina. *Vis. Neurosci.* **16,** 1037–1054 (1999).
13. Rafols, J. A. & Valverde, F. The structure of the dorsal lateral geniculate nucleus of the mouse. A Golgi and electron microscopic study. *J. Comp. Neurol.* **150,** 303–331 (1973).
14. Land, P. W., Kyonka, E. & Shamalla-Hannah, L. Vesicular glutamate transporters in the lateral geniculate nucleus: expression of VGLUT2 by retinal terminals. *Brain Res.* **996,** 251–254 (2004).
15. Kay, J. N. *et al.* Retinal ganglion cells with distinct directional preferences differ in molecular identity, structure, and central projections. *J. Neurosci.* **31,** 7753–7762 (2011).
16. Dhande, O. S. *et al.* Genetic dissection of retinal inputs to brainstem nuclei controlling image stabilization. *J. Neurosci.* **33,** 17797–17813 (2013).
17. Chen, C. & Regehr, W. G. Developmental remodeling of the retinogeniculate synapse. *Neuron* **28,** 955–966 (2000).
18. Huberman, A. D. *et al.* Architecture and activity-mediated refinement of axonal projections from a mosaic of genetically identified retinal ganglion cells. *Neuron* **59,** 425–438 (2008).
19. Völgyi, B., Abrams, J., Paul, D. L. & Bloomfield, S. A. Morphology and tracer coupling patterns of alpha ganglion cells in the mouse retina. *J. Comp. Neurol.* **492,** 66–77 (2005).
20. Chen, T. W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499,** 295–300 (2013).
21. Rivlin-Etzion, M., Wei, W. & Feller, M. B. Visual stimulation reverses the directional preference of direction-selective retinal ganglion cells. *Neuron* **76,** 518–525 (2012).
22. Li, Y. T., Ibrahim, L. A., Liu, B. H., Zhang, L. I. & Tao, H. W. Linear transformation of thalamocortical input by intracortical excitation. *Nature Neurosci.* **16,** 1324–1330 (2013).
23. Lien, A. D. & Scanziani, M. Tuned thalamic excitation is amplified by visual cortical circuits. *Nature Neurosci.* **16,** 1315–1323 (2013).
24. Marshel, J. H., Kaye, A. P., Nauhaus, I. & Callaway, E. M. Anterior-posterior direction opponency in the superficial mouse lateral geniculate nucleus. *Neuron* **76,** 713–720 (2012).
25. Piscopo, D. M., El-Danaf, R. N., Huberman, A. D. & Niell, C. M. Diverse visual features encoded in mouse lateral geniculate nucleus. *J. Neurosci.* **33,** 4642–4656 (2013).
26. Livingstone, M. S. Mechanisms of direction selectivity in macaque V1. *Neuron* **20,** 509–526 (1998).
27. Atallah, B. V., Bruns, W., Carandini, M. & Scanziani, M. Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* **73,** 159–170 (2012).
28. Lee, S. H. *et al.* Activation of specific interneurons improves feature selectivity and visual perception. *Nature* **488,** 379–383 (2012).

Acknowledgements We thank the Kleinfeld laboratory for helpful advice, F. Rieke and M. Turner for example DSGC recording, and the Salk Viral Vector and Biophotonics staff. This work was supported by Vision Core P30 EY019005, the Knights Templar Eye Foundation (O.S.D.), Japan Society for the Promotion of Science (F.O.), Kanae Foundation (F.O.), Uehara Memorial Foundation (F.O.), Naito Foundation (F.O.), NINDS Circuits Training Grant (R.N.E.), Gatsby Charitable Trusts (E.M.C. and A.G.), NIH EY022577 and MH063912 (E.M.C.), Whitehall Foundation (A.D.H.), Ziegler Foundation for the Blind (A.D.H.), Pew Charitable Trusts (A.D.H.), The McKnight Foundation (A.D.H.), and NIH R01EY022157 (A.D.H.).

Author Contributions A.D.H., A.C.-M., A.G. and R.N.E. designed the experiments. A.D.H., A.C.-M., R.N.E. and P.L.N. carried out and analysed the circuit connectivity experiments. A.C.-M. carried out the *in vivo* imaging experiments. B.S. and A.C.-M. analysed imaging data. O.S.D. collected data on molecular markers of cell types. E.M.C. and F.O. designed and made the rabies viruses. A.D.H. and A.C.-M. wrote the paper in collaboration with the other authors. A.D.H. and A.C.-M. prepared the figures. A.D.H. oversaw the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.G. (anirvan.ghosh@roche.com) or A.D.H. (ahuberman@ucsd.edu).

METHODS

All experimental protocols were conducted according to the National Institutes of Health (NIH) guidelines for animal research and were approved by the Institutional Animal Care and Use Committee (IACUC) at the University of California, San Diego, USA.

Mice. Pigmented, wild-type, DRD4-GFP[5], Trhr-GFP[6] mice of both sexes (40 to 60 days old) were used. Animals from different litters and parent mice were used for each experiment. Sample sizes (number of mice and/or cells) for each experiment are stated in main text and figure legends.

Retrograde tracer injection. Mice were anesthetized using an isoflurane–oxygen mixture (4.0% (v/v) for induction and 1.5% (v/v) for maintenance) and given the analgesic buprenorphine (SC, 0.3 mg kg^{-1}). After a midline scalp incision, a small burr hole was made with a pneumatic dental drill (Henry Schein) over V1. For tracer injections into cortex we used either red fluorescent-latex beads (Lumafluor) that travel strictly in retrograde fashion from axon terminals at the injection site to cell bodies of origin[29], or cholera toxin beta (CTβ-594; Invitrogen) which travels both anterograde and retrograde. Tracers (200 nl volume) were injected with a Nanoject II (Drummond) through a small glass pipette (~0.5 MΩ). No distinguishable differences were observable in dLGN-labelling patterns between the two retrograde tracers, except that full-depth injections with CTβ-594 also labelled L6 axons (for example, Fig. 2a, d). All injections were targeted stereotaxically to V1 using the following coordinates: 0.2–0.5 mm anterior to Lambda and 2.5 mm lateral from midline. For superficial injections we penetrated the cortex to a final destination of 150 μm and for deep injections to a depth of 650 μm. Full-depth experiments were accomplished by injecting tracers at multiple cortical depths.

Virus injections. For injections of AAV2 expressing histone-tagged GFP and rabies glycoprotein (AAV2-Glyco-hGFP, titre, 10^{13} genomic titre per ml) into the dLGN a small burr hole was made using the following stereotaxic coordinates: 2.10 mm posterior to Bregma and 2.2 mm lateral from midline. A small glass pipette (~0.5 MΩ) was lowered 2.50 mm from the pial surface and using a Nanoject II (Drummond), 0.04–0.40 μl of virus was inoculated into the brain. A 3-week waiting period was introduced post- AAV2-Glyco-hGFP injection in order to allow maximal expression, after which the scalp was re-incised and a second burr hole was made over V1 for injections of ΔG-RABV-mCherry. It was crucial to inject AAV2-Glyco-hGFP before ΔG-RABV-mCherry because AAVs require more time for expression to occur. ΔG-RABV expressed within 5–7 days and only dLGN neurons expressing both the ΔG-RABV and Glyco-hGFP are competent to infect RGCs that synapse with them. For the cortical injections of ΔG-RABV-mCherry, a pipette (~0.5 MΩ) was lowered to a final destination of either 150 μm ('superficial') or 450 μm ('deep'), and 0.04–0.40 μl of glycoprotein-deleted modified rabies virus encoding mCherry (ΔG-RABV-mCherry, titre, 8.0×10^8–2.0×10^9 infectious units per ml) was inoculated. After surgeries the skin was sutured and the animals were allowed to recover in a heated chamber. For ΔG-RABV-mCherry infections, mice were housed in a biosafety room for 5–7 days to allow rabies virus to infect, trans-synaptically spread, and express mCherry in presynaptic cells.

Virus production. The coding sequence of histone 2B-tagged GFP and rabies virus glycoprotein B19G linked by the F2A element was amplified by PCR with high-fidelity Phusion polymerase[8]. CMV enhancer and synapsin promoter were cloned by PCR[30]. These PCR fragments were inserted into the AAV vector containing woodchuck hepatitis virus posttranscriptional regulatory element. Every plasmid was sequenced before virus production. AAV were generated with transfection of HEK293t cells, purified by iodixanol gradient centrifugation and titred in HEK293t cells with quantitative PCR. The titre of the AAV used in this study was 10^{13} genomic titre per ml (ref. 31). ΔG-RABV expressing mCherry was amplified in B7GG cells, concentrated by two rounds of centrifugation, and titred in HEK293t cells as described previously[8]. The titres of the rabies viruses were 8.0×10^8–2.0×10^9 infectious units per ml. The viruses were stored at −80°C until use.

Retinal histology. Five to seven days after ΔG-RABV-mCherry injection, whole-mount retinas were harvested and fixed in 4% paraformaldehyde (PFA) for 30 min–1 h, then washed in phosphate buffered saline (PBS; 3×, 30 min). The retinas were left for 2 h in blocking solution (10% goat serum, 0.25% Triton-X) at room temperature, then transferred to primary antibodies for 12–18 h at 4 °C, and washed in PBS (3×, 30 min). The retinas were incubated in secondary antibody for 2 h, and washed in PBS (3×, 30 min). Then, retinas were mounted and cover slipped with either Vectashield containing DAPI (Vector Laboratories) or Prolong Gold with DAPI (Invitrogen). Depending on the experiment, the following primary antibodies were used (1:1,000 guinea-pig anti-GFP, Synaptic Systems; 1:1,000 rabbit anti-Cart, Phoenix Pharmaceuticals; 1:1,000 mouse anti-SMI-32, Covance; and 1:1,000 rabbit anti-DSRed, Clontech). The following secondary antibodies were used (1:1,000 goat anti-guinea-pig Alexa Fluor 488; 1:1,000 goat anti-rabbit Alexa Fluor 647; 1:1,000 goat anti-rabbit Alexa Fluor 594; 1:1,000 goat anti-mouse Alexa Fluor 488; Invitrogen).

Retinal physiology. Trhr-GFP mice (6–8 weeks old) were used for targeted recordings of GFP^+ DSGCs using methods described previously[32]. GFP^+ RGCs were identified using two-photon microscopy to minimize photopigment bleaching. Loose patch recording was used to measure spike responses to moving bar stimuli projected onto the photoreceptor layer from an OLED array (eMagin) with mean light levels of approximately 70 rhodopsin isomerizations per rod per second and bar contrasts of 200–300%.

Brain histology. Following retrograde or trans-synaptic tracer injection, mice were deeply anaesthetized with pentobarbital and transcardially perfused with 0.9% saline followed by 4% PFA. Brains were removed, fixed overnight in PFA and then transferred to a 30% (w/v) sucrose solution and stored at 4 °C. Brain slices were collected at 40–60 μm thickness in the coronal plane using a sliding microtome. Brain sections containing dLGN and visual cortex were washed with PBS and incubated for 1 h at room temperature in a blocking solution (10% goat serum, 0.25% Triton-X). This was followed by an overnight incubation with rabbit anti-GFP (1:1,000, Invitrogen) and guinea-pig anti-VGLUT2 (1:1,000, Millipore) primary antibodies in blocking solution at room temperature. Sections were then rinsed with PBS (3×, 30 min), and incubated for 2 h at room temperature with Alexa Fluor 488 goat anti-rabbit (1:1,000, Invitrogen) and Alexa-Fluor 647 goat-anti guinea-pig (1:1,000, Invitrogen) secondary antibody. Lastly, sections were washed with PBS and mounted with Vectashield with DAPI (Vectorlabs).

Targeted filling of genetically tagged RGCs. Targeted RGC filling was conducted using methods described previously[33]. Retinas from mice aged between postnatal day 25–30 (P25-30) mice were dissected and kept in an oxygenated (95% O_2/5% CO_2) solution of Ames' medium (Sigma-Aldrich) supplemented with 23 mM $NaHCO_3$. Borosilicate glass sharp electrodes were used to fill GFP^+ RGCs with 10 mM solution of Alexa 555 hydrazide (Invitrogen) in 200 mM KCl, with the application of hyperpolarizing current pulses ranging between 0.1–0.9 nA for 5–20 min. After cell filling, the retinas were fixed for 1 h with 4% PFA and were processed for immunocytochemistry using the methods described above. The following primary and secondary antibodies were used: rabbit anti-GFP (1:1,000, Invitrogen), guinea-pig anti-VAChT (1:1,000, Millipore), Alexa Fluor 488 goat anti-rabbit and Alexa Fluor 647 goat anti-guinea-pig (1:1,000, Invitrogen). Sections were rinsed with PBS and mounted with Prolong Gold with DAPI (Invitrogen).

Confocal imaging. Fluorescence images were acquired on a laser scanning confocal microscope (Zeiss LSM 710 and 780) equipped with 405, 488, 561 and 633 nm laser lines. Image stacks with a scanning resolution of 1,024 × 1,024 pixels were collected using the following objective lenses: ×10/0.45 Plan-Apochromat (step size of 3.7–4.0 μm), ×20/0.8 Plan-Apochromat (step size of 1.0 μm) and LD C-Apochromat ×40/1.1 water immersion (step size of 0.5–1 μm).

Quantification. For calculating cortical intensity profiles for superficial and deep injections of tracers, the location of the injection was identified from epifluorescent images of tissue sections taken from the entire brain. We evaluated tracer spread along the length of the injection site (from pia to white matter) and the width (parallel to the pial surface). Background fluorescence was subtracted using the contralateral non-injected cortex as the baseline control. After background subtraction, the peak fluorescence along the injection site was quantified by dividing the site into equidistant areas along the injection height and width from pia to the white matter (~800-1,000 μm). For each fluorescence intensity profile, the peak was identified as maximum fluorescence. The fluorescence signal was normalized to this value. Four to five intensity profiles were analysed per tissue section. All tissue sections surrounding the injection site were included in the analysis. We confirmed superficial injections as those with peaks ≤ 150 μm from the pia and deeper injections as those with peaks >300 μm from the pia.

For measuring the location of retrogradely labelled cells along the medial-lateral axis of dLGN, we obtained fluorescent confocal images (×10 and ×20) of coronal brain tissue sections. The boundaries of the dLGN were identified using two stable features: the optic tract, which defines the lateral dLGN border, and the intralaminar zone between the dLGN and the latero-posterior nucleus, which defines the dLGN's medial border. Four measurements were taken per tissue section from locations that spanned the entire rostral-caudal length of the dLGN and thus, the entire retinotopic map. These measurements were placed at equidistant points along the width of the dLGN and an average distance of each labelled cell along the axis from the optic tract (0%) to the medial border (100%) was obtained.

To quantify overlap between retrogradely infected dLGN relay neurons and GFP^+ DSGC axons we obtained fluorescent confocal stacks of coronal tissue sections through the dLGN. For the GFP channel a median filter (2 pixels) was applied and then background subtracted using a rolling ball radius algorithm (50 pixels). After processing images, maximum intensity projections were generated for both channels and the GFP channel was subsequently thresholded using a triangular algorithm[34]. The thresholded image was then binarized to create a mask representing the complete area populated by GFP^+ DSGC axons. This mask was superimposed on the red channel. Every $mCherry^+$ cell body was identified and its

location assessed for overlap with the DSGC-RZ. We note that modified rabies viruses may underestimate total numbers of presynaptic inputs to double-infected cells.

To determine the area of contact between dLGN neurons and VGLUT2 containing DSGC axons, we quantified the percentage of the total somatodendritic area of the mCherry expressing neuron that co-localized with GFP signal from the genetically tagged On-Off DSGC, and that were also immunoreactive for VGLUT2. First, GFP and VGLUT2 puncta were identified by a series of filtering (median = 2 pixels) and erosion steps. This image was then thresholded to identify regions of interest (ROIs) that contained VLGUT2 and GFP puncta, and this was divided by the total measurement of the mCherry labelled area. Images were processed and analysed using algorithms in ImageJ (NIH).

For retinal co-localization experiments, the maximum intensity projections for confocal stacks were obtained for red (mCherry), green (GFP) and far-red channel (Cart). Confocal sections were background subtracted and a maximum projection of the stack was obtained. Channels were merged and markers were considered co-localized if signals from different channels coincided within the same plane.

Statistical analysis. Analysis of fluorescent intensity profiles, normalized distances across the cortex and dLGN and axonal overlap of labelled relay cells and axons was carried out using custom routines in MATLAB (Mathworks) and Image-J (NIH). All statistical analyses were performed with GraphPad Prism (GraphPad Software), and represented as averages. Error bars in graphs represent the s.e.m., where n = number of cells. P value of <0.05 was used to determine significance. All statistical tests were judged as appropriate and described throughout the manuscript. P values for post hoc tests involving multiple comparisons were adjusted using the Tukey's method. No randomization method was used to assign mice to the experimental groups and investigators were blinded to the original group allocation or when assessing an outcome. Variance and normality were tested to ensure that a suitable statistical test was chosen. The variance was similar between groups that were statistically compared. Normality testing in groups with a small number of samples was done with quantile-quantile plots.

For anatomical tracing experiments samples sizes were determined based on preliminary experiments that tested the labelling/infection efficiency of retrograde tracers and/or viral trans-synaptic tools. For imaging experiments no statistical test was used to determine sample size a priori. Experiments were included in the study if they were deemed successful. For the trans-synaptic studies, the experiment was discarded if there were no rabies infected RGCs as assessed by fluorescence. For experiments involving two-photon imaging of axons, data belonging to non-visually responsive axons was not included in the experiment. In addition, mice in which we found no visually driven axons were also excluded from the study.

Imaging of visually evoked calcium signals in DSGC-RZ thalamocortical axons. For monitoring the activity of dLGN axons in V1 either the slow (GCaMP6S, titre, 3.04×10^{13} genomic titre per ml) or the fast (GCaMP6F, titre, 2.96×10^{13} genomic titre per ml) version of the genetically encoded Ca^{2+} sensor GCaMP6 (AAV1.Syn. GCaMP6.WPRE.SV40)[20] were injected into the dLGN using the following coordinates: 2.10 mm posterior to Bregma and 2.2 mm lateral from midline. A small glass pipette (~0.5 MΩ) was then lowered 2.30 mm from the pial surface and using a Nanoject II (Drummond) 0.04-0.20 µl of virus was inoculated into the brain.

Three weeks after the injection of the calcium indicator, mice were deeply anesthetized using an isoflurane–oxygen mixture (4% (v/v) for induction and 1.5% (v/v) for maintenance) and placed in a stereotaxic frame. Following midline incision, a 2 mm diameter craniotomy (1 mm anterior to lambda, 2.5 mm lateral to midline) was performed in the left contralateral primary visual cortex of adult (two month old) C57/Bl6 mice. Next, a sterile 3 mm glass coverslip was gently laid over the dura mater (without using agarose) and glued to the skull with cyanoacrylate glue. Black dental acrylic (Lang Dental) was then applied throughout the skull surface up to the edges of the coverslip and the skin. A custom-made titanium chamber (diameter: 16 mm, thickness: 1.3 mm) was also embedded in the dental acrylic to help secure the mouse onto the stage of the microscope and shield the microscope objective and the photomultipliers from stray light.

Image acquisition and presentation of stimulus grating. Mice were transferred and head-fixed to a stage while under isoflurane anaesthesia and kept at 37 °C using a temperature control device (Harvard Apparatus). Anaesthesia was maintained at 0.8–1.0% (v/v). Imaging was performed using a two-photon microscope system (Sutter) controlled by ScanImage software[35] written in MATLAB (MathWorks). Excitation light from a Mai Tai DeepSee laser (Newport Corp.) with group delay dispersion compensation was scanned by galvanometers (Cambridge Technologies). Images were collected through a ×20 objective (1.0, N.A., Olympus) and the laser power was typically maintained well below 50 mW at the sample. GCaMP6 was excited at 910 nm and emission was collected with a green filter (535 nm centre; 50 nm band; Chroma) via photomultiplier tubes (PMT) (H7422PA-40, Hamamatsu Photonics). Images were collected (×3, digital zoom) at 8 Hz (256 × 128 pixels) or 16 Hz (256 × 64 pixels). For each experiment the optical axis was adjusted to be perpendicular to the cranial window. Imaging was stopped between trials and during this time slow drifts in the optical focus were corrected using a template image. Bleaching of GCaMP6 was not evident during experiments. Because of the limitation of current genetically encoded Ca^{2+} sensors[36] our experiments were not sensitive enough to detect single action potentials, and thus, biased to sample from axons with a high spiking probability.

A visual stimulus containing a full-field two-dimensional moving grating or bar was generated using Psychtoolbox[37,38] in Matlab version 2009. The monitor screen was positioned approximately parallel to, and 29 cm from the right eye of the mouse. The visual angle subtended by the monitor ($\pm 35°$ azimuth $\times \pm 23.5°$ elevation) ensured that projective distortions were small. Both drifting sinusoidal gratings and square wave gratings were used. For sinusoidal gratings, each presentation of a grating started with a grey screen for 3 s, followed by the drifting gratings (0.02 cycles per degree (CPD), 2 Hz, 3 s). In the case of square wave gratings, each trial started with a grey screen for 3 s, followed by the drifting bars (0.01 CPD, 0.33 Hz, 12.75 s). Each trial consisted of presenting eight gratings of different orientations. Within a trial, the orientation of the grating was randomly interleaved. Five to eight trials were recorded to measure average axonal responses to different orientations. Simultaneously, we captured the on-screen luminance by recording the output of a photodiode (Thorlabs) placed at the corner of the screen.

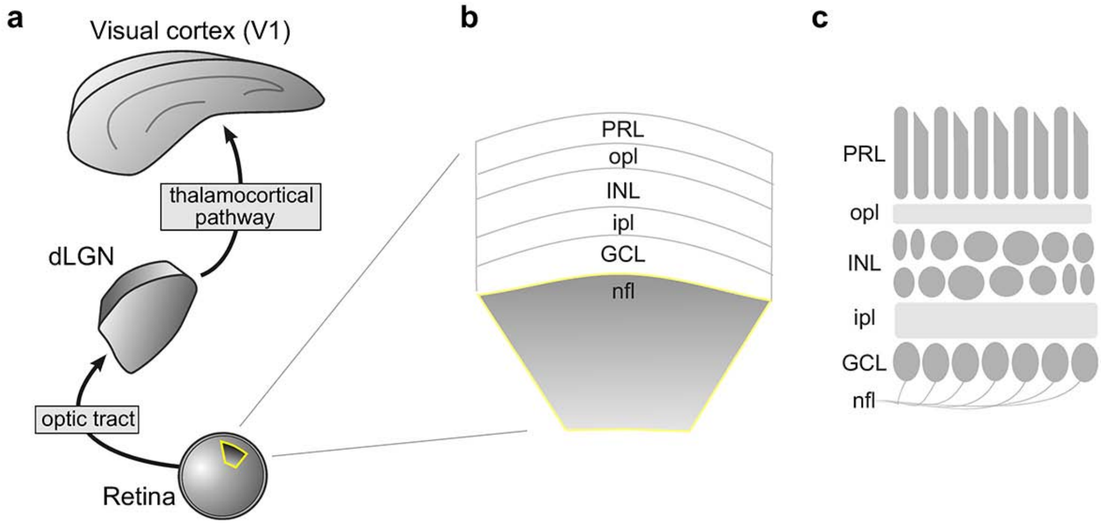
Analysis of time-lapse images and visually evoked fluorescent responses. Frames from time-lapse imaging were registered using template matching (normalized correlation coefficient) and alignment plugins coded in ImageJ (NIH; https://sites.google.com/site/qingzongtseng/template-matching-ij-plugin). To align images we first generated an average image from a small number of frames within a trial where there was minimal movement. A region of interest (ROI) was then selected within this average template image and was used to register all the frames within that repetition.

To extract fluorescence signals, ROIs were drawn over GCaMP6$^+$ varicosities identified by using the mean intensity and standard deviation values of all the repetitions. The pixels in each ROI were averaged to estimate fluorescence corresponding to a single varicosity. Calcium signals were expressed as relative fluorescence changes ($\Delta F/F_0 = ((F - F_0)/F_0)$) corresponding to the mean fluorescence from all pixels within specified ROIs.

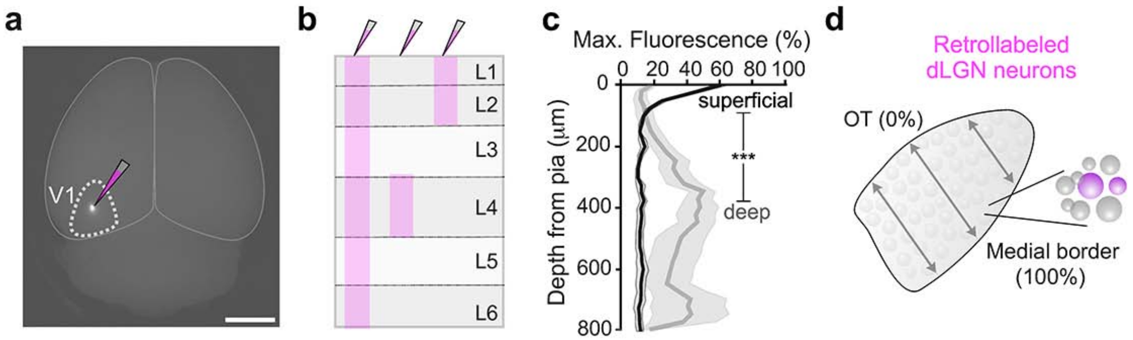
For each experiment, stimulus temporal frequency was measured using the photodiode signal. Putative visually responsive axons were chosen and the calcium signals' response power at the stimulus frequency (F1) and response power at twice stimulus frequency (F2) were measured. To gauge if the measured power was significant, we compared each visually evoked response power with the mean (μ_{grey}) and the standard deviation (σ_{grey}) of calcium response powers to the grey stimulus that immediately preceded each grating. Axons were deemed visually responsive, and included for further analysis, if the F1 power of the visually evoked calcium signal was greater than $\mu_{grey} + 3\sigma_{grey}$ for at least two of the orientations probed. Most (4 out of 5 mice) of the calcium imaging experiments were done using GCaMP6S. For the experiment where we used GCaMP6F we were able to measure frequency doubling responses to the stimulus presentation. Axons were classified as 'linear' if at the orientation corresponding to peak F1 power, the F2 power was smaller than the F1 power. If the F2 power was greater, the axon was judged as 'nonlinear'. In experiments where we used GCaMP6F in combination with lower temporal frequency drifting grating stimulus, we observed two GCaMP6$^+$ varicosities with On-Off responses (Fig. 5i and corresponding polar plot 5l, and polar plot in Fig. 5q; and see imaging methods above). We did not observe differences in the percentage of varicosities that were visually responsive using GCaMP6S compared to GCaMP6F (data not shown); therefore both sets of data are grouped in Fig. 5t, u. Power calculation was performed using the chronux package (mtspectrumc with default tapers; see http://chronux.org/ (ref. 39)). Tuning properties with regard to the orientation of the gratings were quantified by calculating a selectivity index (OSI/DSI). The OSI was defined as $(R_{pref} - R_{ortho})/(R_{pref} + R_{ortho})$, where R_{pref}, the response in the preferred orientation, was the fluorescent response with the peak F1 power. R_{ortho} was similarly calculated as the response evoked by the orthogonal orientation. DSI was defined as $(R_{pref} - R_{opp})/(R_{pref} + R_{opp})$, where R_{opp} is the response in the direction opposite to the preferred direction. Direction tuned neurons were defined as neurons with a selectivity index ≥ 0.4 and orientation selective neurons were defined as those with an index ≥ 0.5.

29. Katz, L. C., Burkhalter, A. & Dreyer, W. J. Fluorescent latex microspheres as a retrograde neuronal marker for *in vivo* and *in vitro* studies of visual cortex. *Nature* **310,** 498–500 (1984).
30. Hioki, H. *et al.* Efficient gene transduction of neurons by lentivirus with enhanced neuron-specific promoters. *Gene Ther.* **14,** 872–882 (2007).
31. Osakada, F. & Callaway, E. M. Design and generation of recombinant rabies virus vectors. *Nature Protocols* **8,** 1583–1601 (2013).
32. Murphy, G. J. & Rieke, F. Network variability limits stimulus-evoked spike timing precision in retinal ganglion cells. *Neuron* **52,** 511–524 (2006).

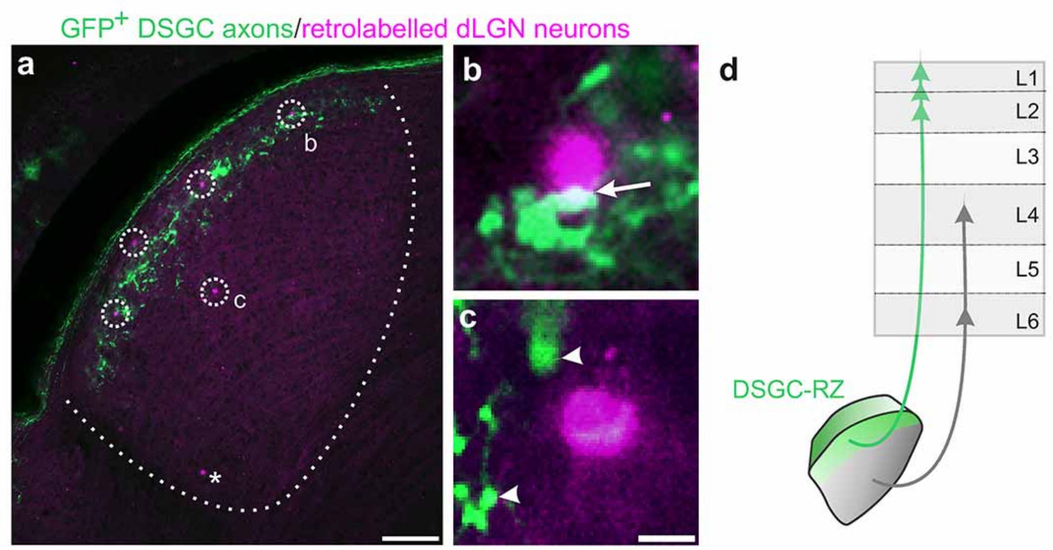
33. Beier, K. T. *et al.* Transsynaptic tracing with vesicular stomatitis virus reveals novel retinal circuitry. *J. Neurosci.* **33,** 35–51 (2013).
34. Zack, G. W., Rogers, W. E. & Latt, S. A. Automatic measurement of sister chromatid exchange frequency. *J. Histochem. Cytochem.* **25,** 741–753 (1977).
35. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2,** 13 (2003).
36. Tian, L. *et al.* Imaging neural activity in worms, flies and mice with improved GCaMP indicators. *Nature Methods* **6,** 875–881 (2009).
37. Brainard, D. H. The psychophysics toolbox. *Spat. Vis.* **10,** 433–436 (1997).
38. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10,** 437–442 (1997).
39. Bokil, H., Andrews, P., Kulkarni, J. E., Mehta, S. & Mitra, P. P. Chronux: a platform for analyzing neural signals. *J. Neurosci. Methods* **192,** 146–151 (2010).



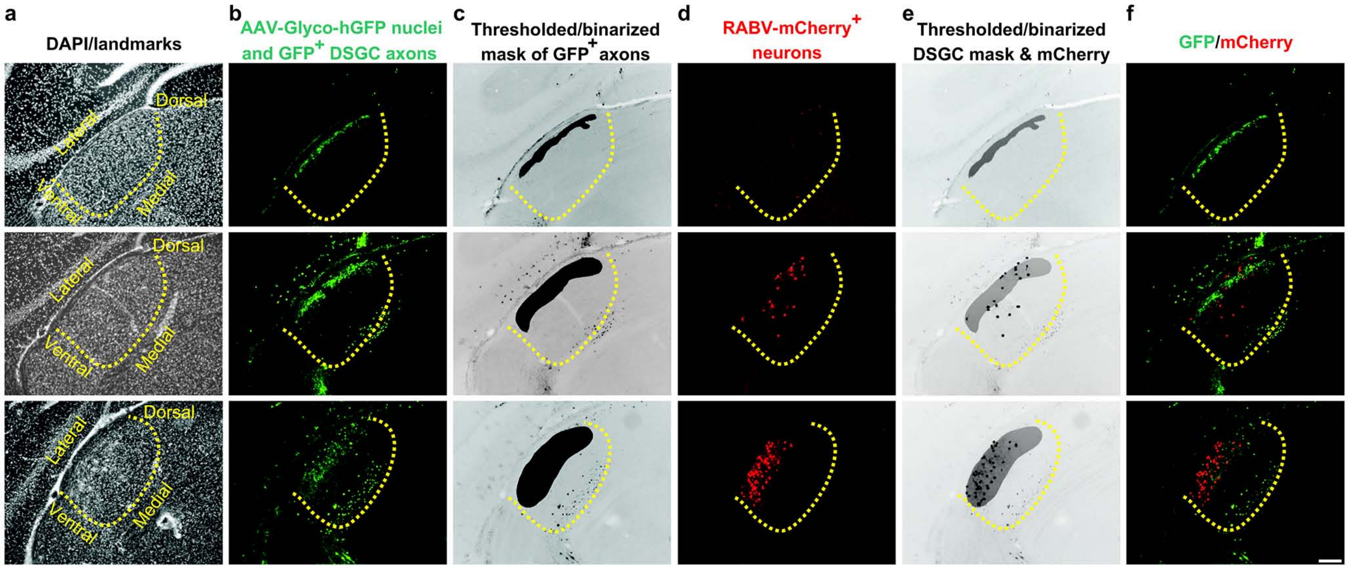
Extended Data Figure 1 | The retino-geniculo-cortical pathway links retinal cells and circuits to the brain. **a**, Diagram of retina, dorsal lateral geniculate nucleus (dLGN) and primary visual cortex (V1). The optic tract which carries retinal ganglion cell (RGC) axons and thalamocortical (dLGN to V1) pathway also shown. **b**, Diagram of retinal layers: PRL, photoreceptor layer; opl, outer plexiform layer; INL, inner nuclear layer; ipl, inner plexiform layer; GCL, ganglion cell layer; nfl, nerve fibre layer. **c**, Retina diagram with cells shown (labels same as in **b**).



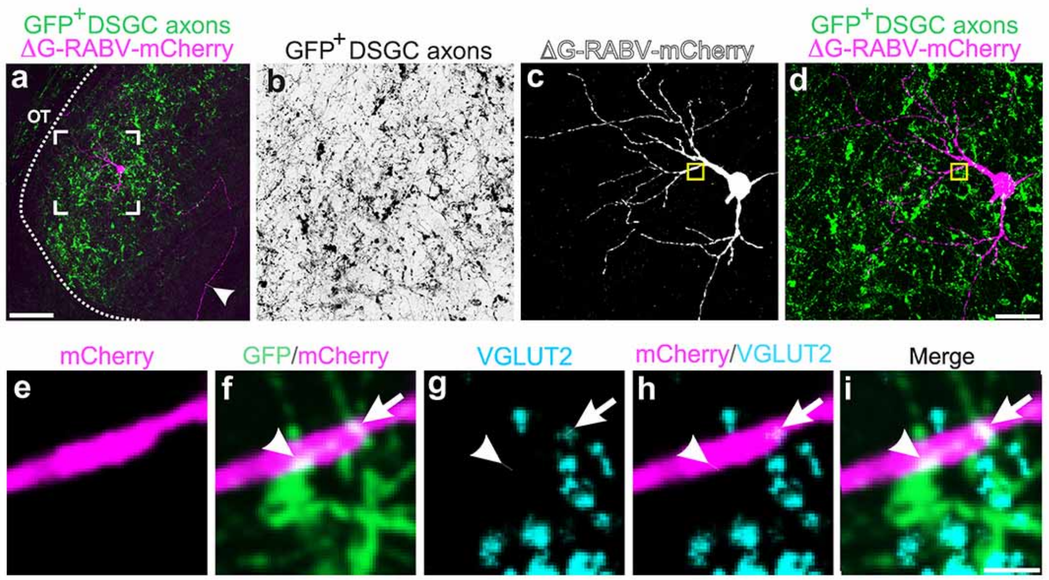
Extended Data Figure 2 | Approach for assessing laminar specificity of mouse geniculocortical projections. **a**, Focal retrograde tracer injection to V1. Scale bar, 3 mm. **b**, Diagram of the three different injection depths used to generate data in Fig. 2. **c**, Percentage of fluorescence in V1 from superficial (black line) versus deep (grey line) injections. Superficial, peak intensity occurs at 25 μm from pial surface (4 mice). Deep, peak intensity occurs at 350 μm from pial surface. Gray shaded regions, s.e.m. (superficial vs deep = ***$P < 0.0001$; two-way ANOVA). **d**, Assessment of retrogradely labelled cells across the width of the dLGN. 0% is at optic tract, 100% is at medial border (see Fig. 2g–i).



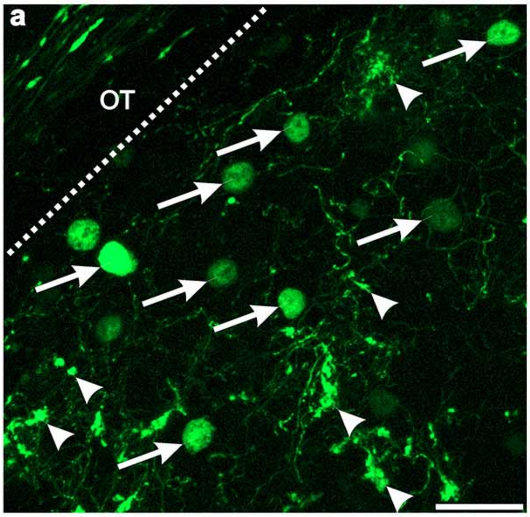
Extended Data Figure 3 | Retrograde tracers to superficial V1 label cells in the DSGC-RZ. a–c, Same dLGN as in main Fig. 2f but with GFP$^+$ On-Off DSGC[6] axons shown. **a**, most of the retrogradely labelled cells (magenta/dashed circles) reside in the DSGC-RZ (green terminals). Asterisk, labelled cell outside the DSGC-RZ. Scale bar, 200 μm. **b**, **c**, High magnification views of retrogradely labelled dLGN neuron cell bodies with potential contact from GFP$^+$ DSGC axons (arrow in **b**); **c**, this cell is in vicinity of DSGC axonal boutons (arrowheads). **b**, **c**, Scale, 15 μm. **d**, Diagram of laminar-specific connections between DSGC-RZ and superficial V1 and dLGN core and deeper V1 layers 4 and 6.



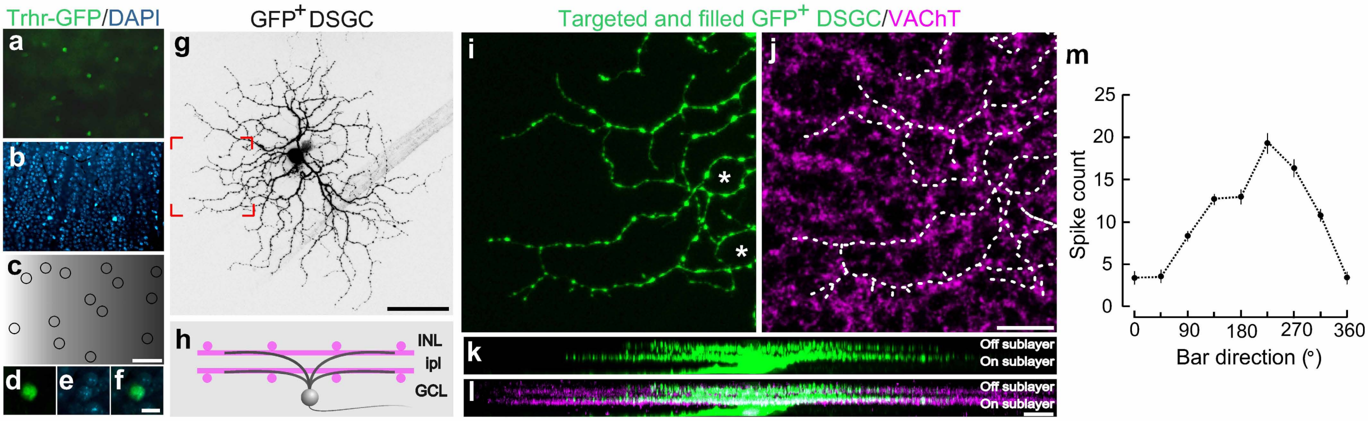
Extended Data Figure 4 | Analysis of dLGN neurons retrogradely infected from superficial V1. **a**–**f**, Example serial sections of anterior, middle and posterior portions of dLGN in a mouse with GFP expressing On-Off DSGC axons that was injected with ΔG-RABV-mCherry in superficial layers of V1. **a**, DAPI to show cytoarchitectural landmarks and dLGN borders. **b**, GFP^{+} DSGC axons and AAV2-Glyco-hGFP-infected cell bodies (see main Fig. 4 and text). **c**, Mask of GFP^{+} DSGC axons (Methods). **d**, ΔG-RABV-$mCherry^{+}$ dLGN relay neurons. **e**, GFP^{+} DSGC axon mask superimposed with mCherry signal; this was used to determine colocalization. **f**, mCherry and GFP signals merged. Scale bar, 200 μm.



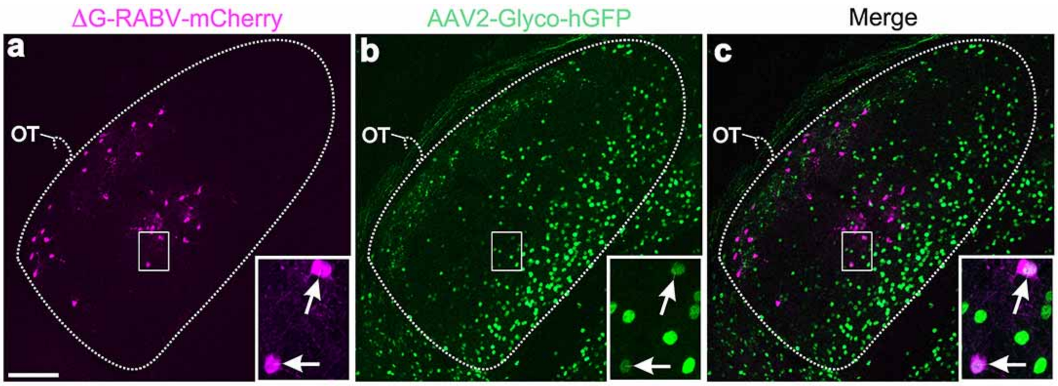
Extended Data Figure 5 | Putative sites of contact between DSGC axons and a dLGN neuron retrogradely infected from superficial V1. a–i, GFP$^+$ On-Off DSGC axons (green in all panels except black in **b**) and mCherry$^+$ dLGN relay neuron (magenta in all panels except white in **c**) infected by injection to superficial V1. Framed region in **a** is shown at higher magnification in **b–d**. Arrowhead (**a**), thalamocortical axon of mCherry$^+$ dLGN cell. Scale bar in **a**, 50 μm. Yellow boxed region in **c**, **d**, is shown at higher magnification in **e–i**. Scale bar in **d**, 15 μm. **e–i**, Some DSGC axon–dendrite contacts contain VGLUT2 (blue). **f–i**, Arrowhead, site of GFP/mCherry co-localization that does not contain VGLUT2; arrow, GFP/mCherry/VGLUT2$^+$ contact.



Extended Data Figure 6 | The axons of GFP^{+} On-Off DSGCs and dLGN neurons infected with AAV2-Glyco-hGFP can be distinguished on the basis of their cellular localization. High magnification view of DSGC-RZ in mouse with GFP^{+} posterior-tuned On-Off DSGCs that was injected 14 days earlier with AAV2-Glyco-hGFP. Glyco-hGFP^{+} neurons have nuclear GFP labelling (arrows), whereas DSGCs have GFP in axon terminals (arrowheads). Dashed line, lateral border of dLGN. OT, optic tract. Scale bar, 50 μm.



Extended Data Figure 7 | Signature anatomical and physiological characteristics of GFP-tagged On-Off DSGCs. **a**, **b**, Flat-mount retina with GFP^{+} On-Off DSGCs (**a**) and co-stained with DAPI (**b**). **c**, Positions of GFP^{+} RGCs. Scale bar in **c**, 150 µm. **d**–**f**, High magnification views. Scale bar, 12 µm. **g**, Targeted fill of a GFP^{+} DSGC. Scale bar, 50 µm. **h**, Schematic of On-Off DSGC stratification and starburst amacrine cells (magenta). Labelling as in Extended Data Fig. 1. **i**, **j**, Higher magnification of framed region in **g** stained for VAChT (starburst amacrine processes). Asterisk, 'looping arborizations'; dashed line, GFP arborization, which matches VAChT plexus. Scale bar, 10 µm. **k**, **l**, Side (x–z plane) views of cell in **g**. GFP^{+} dendrites co-stratify with both the On and Off sublayers. Scale bar, 5 µm. **m**, Direction-tuned response of a GFP^{+} On-Off DSGC targeted for recording and receptive field characterization. The spike count is highest for bars moving towards ~270° in the cardinal axes.



Extended Data Figure 8 | Injections of ΔG-RABV-mCherry into both superficial and deep V1 combined with AAV2-Glyco-hGFP infection of dLGN core. **a**, mCherry^{+} neurons in the DSGC-RZ and the core of the dLGN. **b**, AAV2-Glyco-hGFP: many neurons throughout the dLGN, but mostly along the medial border and not in the shell/DSGC-RZ express Glyco-hGFP. DSGC-RZ marked by axons of GFP^{+} On-Off DSGCs. **c**, Merged of **a**, **b**. Scale in **a**, 100 µm. Boxed regions with arrows: two dLGN neurons; both RABV-mCherry^{+} and AAV2-Glyco-hGFP^{+}. One or both of these cells infected their presynaptic partner, the RGC shown in Fig. 4 (panels cc-ee) of the main text. Scale bar, 15µm.

LETTER

doi:10.1038/nature12972

Intestinal crypt homeostasis revealed at single-stem-cell level by *in vivo* live imaging

Laila Ritsma[1]*, Saskia I. J. Ellenbroek[1]*, Anoek Zomer[1], Hugo J. Snippert[2], Frederic J. de Sauvage[3], Benjamin D. Simons[4,5,6], Hans Clevers[1] & Jacco van Rheenen[1]

The rapid turnover of the mammalian intestinal epithelium is supported by stem cells located around the base of the crypt[1]. In addition to the Lgr5 marker, intestinal stem cells have been associated with other markers that are expressed heterogeneously within the crypt base region[1–6]. Previous quantitative clonal fate analyses have led to the proposal that homeostasis occurs as the consequence of neutral competition between dividing stem cells[7–9]. However, the short-term behaviour of individual Lgr5$^+$ cells positioned at different locations within the crypt base compartment has not been resolved. Here we establish the short-term dynamics of intestinal stem cells using the novel approach of continuous intravital imaging of Lgr5-Confetti mice. We find that Lgr5$^+$ cells in the upper part of the niche (termed 'border cells') can be passively displaced into the transit-amplifying domain, after the division of proximate cells, implying that the determination of stem-cell fate can be uncoupled from division. Through quantitative analysis of individual clonal lineages, we show that stem cells at the crypt base, termed 'central cells', experience a survival advantage over border stem cells. However, through the transfer of stem cells between the border and central regions, all Lgr5$^+$ cells are endowed with long-term self-renewal potential. These findings establish a novel paradigm for stem-cell maintenance in which a dynamically heterogeneous cell population is able to function long term as a single stem-cell pool.

In the small intestine, stem cells are associated with Lgr5 expression, which marks around 14–16 proliferative crypt base columnar (CBC) cells distributed throughout the crypt base. The stem-cell niche is constituted by Paneth cells[10,11] and the surrounding mesenchyme[12]. Cells that become displaced from this region enter the transit-amplifying domain (TA) and lose stemness[13]. Quiescent or slow-cycling cells, positioned at or near the +4 position may constitute a second stem-cell type[3,5,6,14], although a recent study indicated that some, if not all, of these cells represent secretory precursors that, in common with Dll1$^+$ cells higher in the crypt[15], can be recruited back into the stem-cell compartment upon damage[16]. The hierarchy, heterogeneity and spatial organization of intestinal stem cells remain subjects of debate[17–21]. It is unclear whether stem and progenitor cells are organized in an engrained proliferative hierarchy, defined by the signature of molecular markers, or whether stem cells transit reversibly between states of variable competence in which they become biased towards renewal or differentiation. If the latter is true, then it remains to be determined whether that bias is controlled by intrinsic heterogeneity in the expression of fate determinants, or is the consequence of spatiotemporal cues associated with niche-derived signals. Although inducible genetic lineage tracing allows us to dissect short-term heterogeneity in self-renewal potential, its reliability may be undermined by transient effects due to drug-inducing agents, Cre activity, or non-representativeness of labelling[22]. Therefore we applied an *in vivo* live-imaging strategy, allowing measurements to begin several days after drug administration. In common with previous live-imaging approaches used to study stem cells in hair follicle and testis[23–25], our approach enables tracing of the fate of individual marked stem cells and their progeny over time *in vivo*.

Multiphoton intravital microscopy and surgical implantation of an abdominal imaging window (AIW)[26,27] into living Lgr5$^{\text{eGFP-Ires-CreERT2}}$/R26R-Confetti mice were used to obtain visual access to the intestinal stem-cell niche (Fig. 1a). Lgr5$^+$ CBC cells and their progeny were lineage traced over time (Extended Data Fig. 1) by activating the expression of one of the Confetti colours (membranous cyan fluorescent protein (CFP), cytoplasmic yellow fluorescent protein (YFP) and red fluorescent protein (RFP)) in individual Lgr5$^+$ cells using tamoxifen-mediated recombination of the Confetti construct (Fig. 1a). To characterize the fate behaviour of CBC cells, we followed lineages of 80 marked cells ($n = 4$ mice) up to 5 days from the start of time-lapse imaging (Extended Data Fig. 2; for controls, see ref. 27 and Extended Data Fig. 3).

After induction, clonal progeny were observed throughout the stem-cell niche. To quantify the fate behaviour of Lgr5$^+$ CBC cells, we acquired z-stacks (Fig. 1b; see Supplementary Video 1 for the three-dimensional reconstruction) and classified cells based on their relative position, using the most basal cells (row 0) as a reference (Fig. 1b). Confetti-labelled clones were scored according to cell number, disaggregated by position (Extended Data Fig. 4). In line with predictions of neutral competition[7], numbers of marked cells in the stem-cell niche varied widely between clones (some expanded in size, others lost attachment to this compartment altogether; Extended Data Figs 2, 4). As just 1 of the 28 clones containing a single marked CBC cell at the start of filming remained single after 2 days of tracing, we chose to neglect the potential impact of lineage-committed quiescent Lgr5$^+$ cells, which was identified previously[16].

To investigate spatial heterogeneity in the self-renewal potential of CBC cells, we defined two regions within the Lgr5$^+$ stem-cell niche: a central (rows 0 to +2) and border (+3 and +4) region (Fig. 1b). A 'mother' cell in either the central or the border region could expand and give rise to progeny that extended into both regions (Fig. 1c–f and Extended Data Fig. 5). Further quantitative analysis was necessary to address the potency of CBC cells in these two domains. Although the average number of central cells per clone derived from a single central mother cell remained approximately constant, consistent with their maintenance over time, the average number of border cells derived from these cells increased to approximately two by day 3 (Fig. 2a). Furthermore, maintenance of the average central cell number was achieved through the steady decline in the number of clones retaining at least one central cell (Fig. 2b), compensated by a steady increase in size of those that remained (Fig. 2c). Although clones derived from single border mother cells also appeared to approximately maintain their number, they gave rise to a comparatively smaller number of central cells (Fig. 2a). The sustained increase in the number of border cells from a central

[1]Cancer Genomics Netherlands, Hubrecht Institute-KNAW and University Medical Centre Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands. [2]University Medical Centre Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands. [3]Department of Molecular Biology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. [4]Cavendish Laboratory, Department of Physics, J. J. Thomson Avenue, University of Cambridge, Cambridge CB3 0HE, UK. [5]The Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK. [6]The Wellcome Trust/Medical Research Council Stem Cell Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK.
*These authors contributed equally to this work.

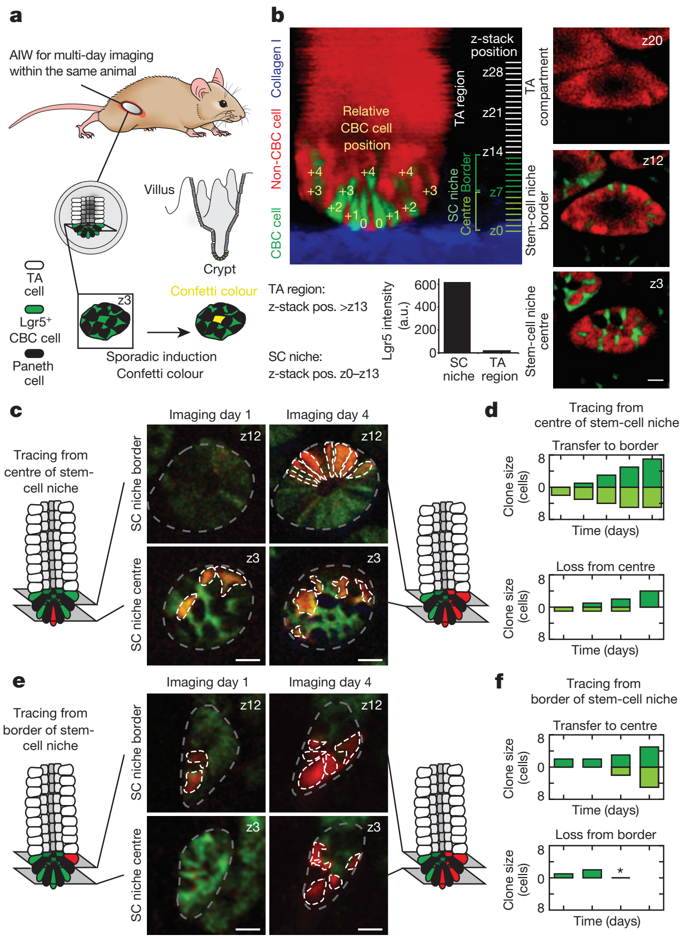


Figure 1 | Intravital lineage tracing of Lgr5$^+$ cells. a, Cartoon showing a Lgr5$^{eGFP\text{-}Ires\text{-}CreERT2}$/R26R-Confetti mouse with an AIW to visualize intestinal Lgr5$^+$ CBC cells and their Confetti progeny over multiple imaging sessions. **b**, Lateral projection of a z-stack and representative *xy* images of a crypt at indicated z-stack positions (pos.). The stem-cell (SC) niche (z0–13) is defined by Lgr5–GFP fluorescence. The relative position of CBC cells to the most basal cell (row 0) determines location in the central (row 0 to +2, which translates to z0–6) or border region (row +3 to +4, which translates to z7–13) of the stem-cell niche. a.u., arbitrary units. Scale bar, 20 µm. **c–f**, Intravital lineage tracing of RFP-expressing Lgr5$^+$ CBC cells located at the centre (**c**, **d**) and border (**e**, **f**) regions. Grey lines indicate crypts, white lines indicate Confetti clones. **d**, **f**, Graphs show time evolution of spatial organization of Confetti clones starting 3 days after induction. Clone size is divided into central (light green) and border (dark green) CBC cells. Asterisk indicates clones in which all progeny were lost. Scale bar, 20 µm.

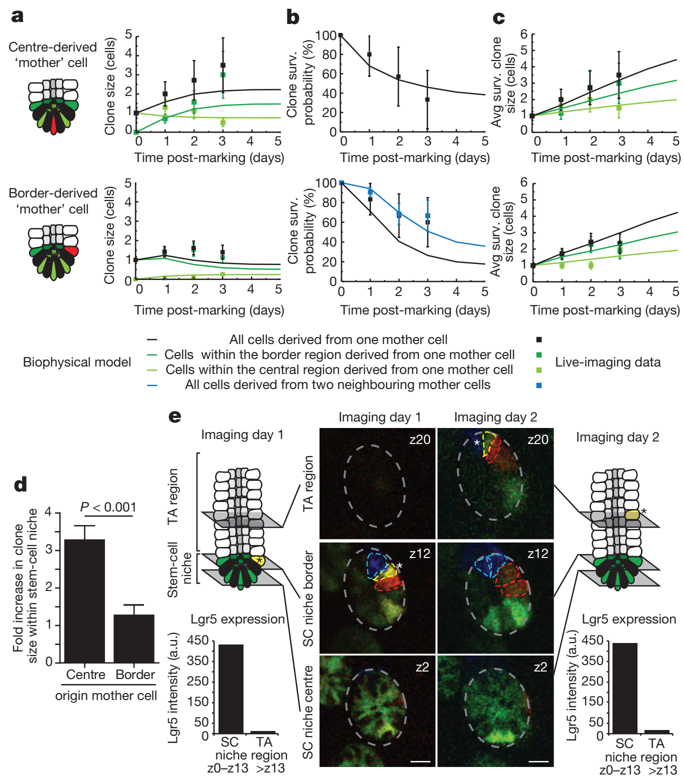


Figure 2 | Central CBC cells experience a short-term positional advantage in self-renewal potential. a–c, Clonal evolution of a Confetti cell located at the central or border region starting 3 days after induction. Graphs show average clone size (**a**), fraction of 'surviving' clones that contain at least one marked central (top) or border (bottom) cell (**b**), and average size of surviving clones (clones with at least one marked cell) (**c**). Different colours indicate different regions in the niche. Points show data and lines show fit to the biophysical model (see Fig. 3). Error bars represent standard deviation (s.d.). **d**, Fold increase in clone size over 3 days from a border or central Confetti$^+$ CBC cell. Note that central stem cells have a positional advantage over border stem cells. Error bars represent standard error of the mean (s.e.m.); $P < 0.001$ obtained using a Mann–Whitney U test. **e**, Intravital images of the same crypt at indicated times. Note that the yellow cell is truly expelled from the stem-cell (SC) niche, as GFP expression was absent in the TA cell region (see charts at indicated time points). Scale bars, 20 µm.

mother cell (Fig. 2a) indicates that these cells typically outcompete cells at the niche border (Fig. 2d).

To investigate the potential basis of this positional advantage, we studied the development of clones with finer time resolution. Every 2 hours, we acquired multiphoton images of crypts, followed the location of all green fluorescent protein (GFP)-labelled cells over time (Extended Data Fig. 6 and Supplementary Videos 2–4), and found that division of single Lgr5$^+$ cells coincides with displacement of proximate CBC cells. This suggests that cell proliferation creates competition for space, leading to an adjustment of cell positions. Through this rearrangement, and independent of their division history, CBC cells located at the border can become passively displaced from the niche after division of a neighbour (Fig. 2e and Extended Data Fig. 4).

To challenge this conclusion and address the potency of the Lgr5$^+$ CBC stem-cell population, we aimed to capture quantitatively the variability seen in the lineage potential of individual cells by a biophysical model, involving a revision of the neutral drift dynamics model introduced previously[7,8], in which all stem cells were considered functionally equivalent. In this new model, a periodic quasi-one-dimensional arrangement of stem cells mimicked the 'collar-like' geometry of the central and border niche regions of the crypt (Fig. 3a). To account for the mixed GFP expression profile seen at rows +3 and +4 (Fig. 1b), the border region was further subdivided into Lgr5$^+$ CBC cells and Lgr5$^-$ TA cells. To accommodate the range of observed dynamical behaviours in the stem-cell niche, we allowed for five possible 'channels' of stem-cell loss and replacement (Fig. 3a). That is, after division of a border stem cell, one daughter cell remains at its position while the other either (1) displaces a border TA cell out of the niche; (2) displaces a border stem cell that in turn displaces a border TA cell out of the niche; or (3) displaces a central stem cell that in turn displaces a border stem cell into the border TA cell domain. Similarly, after division of a central stem cell, one daughter remains at its position while the other either (4) displaces a border stem cell into the border TA cell region; or (5) displaces a central cell that in turn displaces a border stem cell into the TA cell region. If we define λ as the rate of transfer of border TA cells out of the niche, each of these five processes occur at rates $P_b\lambda$, $P_{bb}\lambda$, $P_{bc}\lambda$, $P_{cb}\lambda$ and $P_{cc}\lambda$, respectively, with $P_b + P_{bb} + P_{bc} + P_{cb} + P_{cc} = 1$ (Supplementary Notes).

By fixing the relative rates of stem-cell division and displacement by the observed average clone size dependences and independent estimates

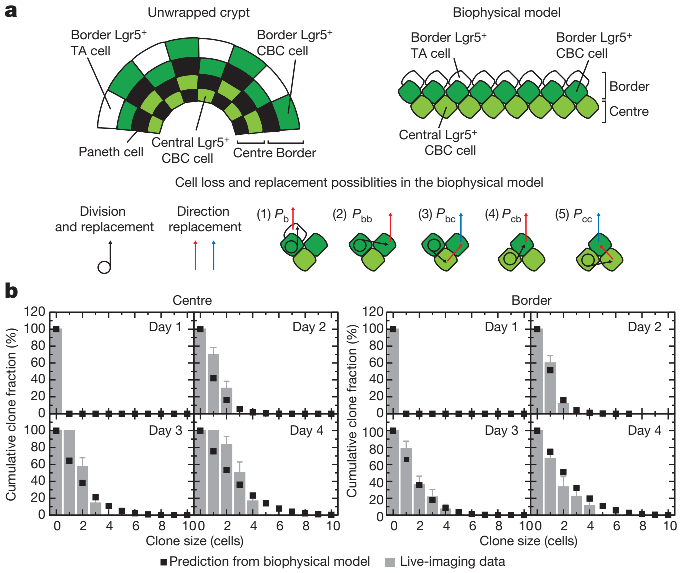


Figure 3 | Biophysical model of intestinal stem-cell dynamics. **a**, From the unfolded crypt caricature (left), we synthesize a quasi-one-dimensional biophysical model of the niche region (right) consisting of two domains: border and centre. To conserve cell number, cell rearrangements after stem-cell division displace precisely one cell from the border. To capture the range of lineage data, we include five channels of stem-cell loss/replacement (1–5), as defined in the main text. **b**, Cumulative size distributions of clones derived from a single cell in the centre (left) or border (right). Clone size is defined in both cases by total number of constituent cells in the centre and border regions. Error bars represent s.e.m. Points represent predictions of the model using the same parameters as those inferred from the average dependences (Supplementary Notes).

of the average cell division rate, we found that the biophysical model can accurately predict the clone size distribution and spatial dependencies observed in live imaging (Figs 2a–c, 3b and Extended Data Fig. 7). More significantly, with the same parameters, the model describes quantitatively convergence onto the hallmark scaling behaviour reported using static lineage tracing assays at intermediate times[7] (7 and 14 days after induction), as well as the predicted progression towards crypt monoclonality at long times[8] (Extended Data Figs 8, 9 and Supplementary Notes).

To challenge the model further, we traced the recovery of stem cells after targeted ablation of Lgr5$^+$ cells using diphtheria toxin (DT) injection in mice in which the human DT receptor (DTR) fused to enhanced (e)GFP was knocked in at the *Lgr5* locus (Lgr5$^{\text{DTR:eGFP}}$)[28] (Fig. 4a). In these mice, recovered Lgr5$^+$ cells are derived from a TA lineage[28]. After complete depletion (Fig. 4a), we observed a low frequency of initiation and a heterogeneous pattern of recovery (Fig. 4b and Supplementary Video 5), suggesting sporadic transfer of cells from the TA zone into the stem-cell niche border. The cohesion of these recovered cell clusters (Supplementary Video 6) suggests clonal expansion of individual TA cells. Intriguingly, by allowing individual border stem cells to recolonize a depleted stem-cell niche through cell division uncompensated by loss, our biophysical model provided a quantitative prediction of cluster composition (border versus central) by size, with the same relative rates of stem-cell division as those found in the steady state (Fig. 4c and Supplementary Notes).

Our data show that intestinal stem-cell maintenance follows from competition between proximate CBC stem cells for limited niche access, and that stem cells positioned near the niche boundary experience a bias towards loss and replacement, whereas stem cells remote from the boundary are biased towards survival. Intriguingly, a similar dependence of self-renewal potential on proximity to the niche border was reported in a recent *in vivo* live-imaging study of the mouse hair follicle[29], suggesting that such heterogeneity may be a ubiquitous feature of adult stem-cell populations. A recent lineage tracing study based on the continuous and sporadic acquisition of mutations during DNA replication concluded that only a subfraction of putative intestinal stem cells are 'functional'[30]. Our quantitative analysis of live-imaging data shows that central stem cells are about three times more likely than border cells to colonize fully a crypt in steady state, explaining why only a fraction of Lgr5$^+$ cells seem to retain long-term self-renewal

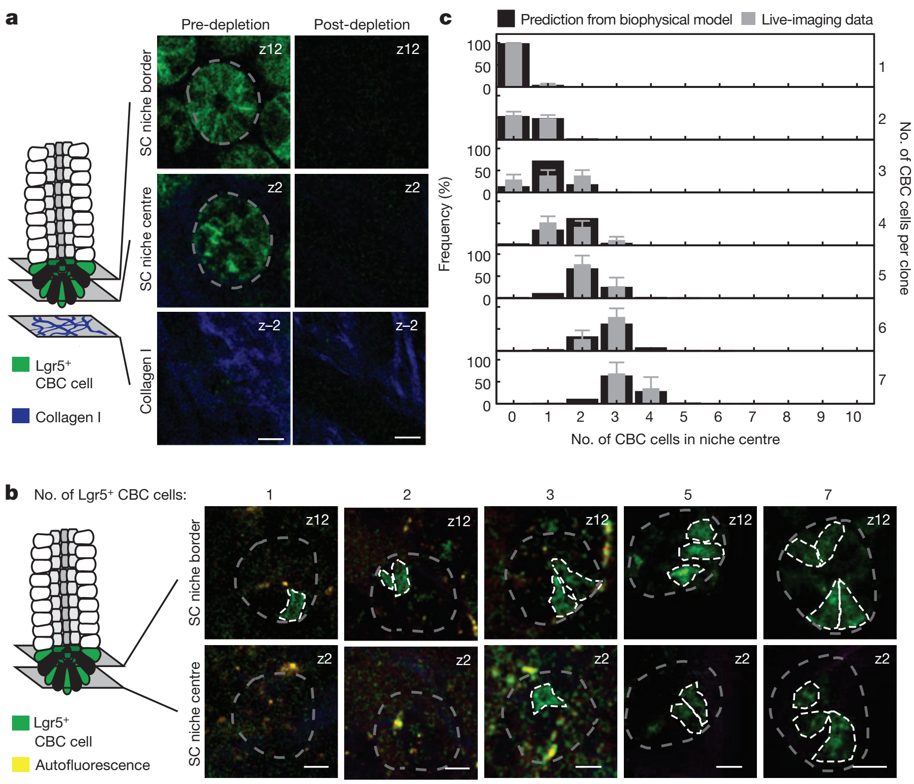


Figure 4 | Recovery of stem-cell compartment after ablation of Lgr5$^+$ cells challenges model. **a**, Targeted ablation of Lgr5$^+$ cells in Lgr5$^{\text{DTR:eGFP}}$ mice was induced by injection of DT. Shown are representative images pre- and post-ablation. Scale bars, 20 µm. **b**, Recovery of Lgr5$^+$ CBC cells was monitored only in mice in which full depletion was confirmed 24 h after DT injection. Images taken at 72 h after depletion show representative crypts containing clonal clusters of different sizes ($n = 108$ crypts in 3 mice). Scale bars, 20 µm. **c**, For all various clone sizes, measured spatial composition (border versus centre) of Lgr5$^+$ CBC cells in clusters (grey) was accurately predicted by the biophysical model (black). Error bars represent s.d.

potential (Supplementary Notes). Through the transfer of cells between the central and border regions of the niche, the dynamic and heterogeneous population of intestinal stem cells is able to function long term as a single equipotent pool.

METHODS SUMMARY

All experiments were carried out in accordance with the guidelines of the Animal Welfare Committee of the Royal Netherlands Academy of Arts and Sciences. Lgr5$^{eGFP\text{-}Ires\text{-}CreERT2}$/R26R-Confetti mice between 10 and 22 weeks old were used for experiments. Three days before imaging, mice were injected with 2.5–5 mg tamoxifen (Sigma Aldrich) to activate expression of Confetti colours. Lgr5$^{DTR:eGFP}$ mice with an AIW received 50 μg kg^{-1} DT through intraperitoneal injections. Intravital imaging was performed on an inverted Leica TCS SP5 AOBS two-photon microscope with a chameleon Ti:Sapphire pumped optical parametric oscillator (Coherent) equipped with a ×25 (HCX IRAPO NA0.95 WD 2.5 mm) water objective and four non-descanned detectors (NDDs). Confetti colours were detected using sequential scanning: excitation 860 nm, NDD2 455–490 nm emission (CFP and eGFP); excitation 960 nm, NDD3 500–550 nm emission (eGFP and YFP), NDD4 560–650 nm emission (YFP and RFP). Second harmonic generation (SHG) signal was generated by 960 nm excitation at collagen I and detected in NDD2. z-stacks with 2.5 μm z-steps of typically 70–80 images were acquired. Re-identification of the same crypts over multiple days was accomplished by storing the *xy* coordinates of the imaged regions using the 'multiple position' function in the LAS-AF software and using the vasculature and the typical Lgr5^{+} crypt pattern as visual landmarks. Acquired images were processed using a custom-designed Visual Basic program and further processed with ImageJ (National Institutes of Health).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 May; accepted 20 December 2013.
Published online 16 February 2014.

1. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449,** 1003–1007 (2007).
2. Sangiorgi, E. & Capecchi, M. R. *Bmi1* is expressed *in vivo* in intestinal stem cells. *Nature Genet.* **40,** 915–920 (2008).
3. Takeda, N. *et al.* Interconversion between intestinal stem cell populations in distinct niches. *Science* **334,** 1420–1424 (2011).
4. Montgomery, R. K. *et al.* Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells. *Proc. Natl Acad. Sci. USA* **108,** 179–184 (2011).
5. Powell, A. E. *et al.* The Pan-ErbB negative regulator Lrig1 is an intestinal stem cell marker that functions as a tumor suppressor. *Cell* **149,** 146–158 (2012).
6. Wong, V. W. Y. *et al.* Lrig1 controls intestinal stem-cell homeostasis by negative regulation of ErbB signalling. *Nature Cell Biol.* **14,** 401–408 (2012).
7. Snippert, H. J. *et al.* Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143,** 134–144 (2010).
8. Lopez-Garcia, C., Klein, A. M., Simons, B. D. & Winton, D. J. Intestinal stem cell replacement follows a pattern of neutral drift. *Science* **330,** 822–825 (2010).
9. Snippert, H. J. & Clevers, H. Tracking adult stem cells. *EMBO Rep.* **12,** 113–122 (2011).
10. Sato, T. *et al.* Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469,** 415–418 (2011).
11. VanDussen, K. L. *et al.* Notch signaling modulates proliferation and differentiation of intestinal crypt base columnar stem cells. *Development* **139,** 488–497 (2012).
12. Farin, H. F., Van Es, J. H. & Clevers, H. Redundant sources of Wnt regulate intestinal stem cells and promote formation of Paneth cells. *Gastroenterology* **143,** 1518–1529 (2012).
13. van der Flier, L. G. & Clevers, H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annu. Rev. Physiol.* **71,** 241–260 (2009).
14. Sangiorgi, E. & Capecchi, M. R. Bmi1 lineage tracing identifies a self-renewing pancreatic acinar cell subpopulation capable of maintaining pancreatic organ homeostasis. *Proc. Natl Acad. Sci. USA* **106,** 7101–7106 (2009).
15. van Es, J. H. *et al.* Dll1^{+} secretory progenitor cells revert to stem cells upon crypt damage. *Nature Cell Biol.* **14,** 1099–1104 (2012).
16. Buczacki, S. J. A. *et al.* Intestinal label-retaining cells are secretory precursors expressing Lgr5. *Nature* **495,** 65–69 (2013).
17. Stine, R. R. & Matunis, E. L. Stem cell competition: finding balance in the niche. *Trends Cell Biol.* **23,** 357–364 (2013).
18. Simons, B. D. & Clevers, H. Strategies for homeostatic stem cell self-renewal in adult tissues. *Cell* **145,** 851–862 (2011).
19. Morrison, S. J. & Spradling, A. C. Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. *Cell* **132,** 598–611 (2008).
20. Goulas, S., Conder, R. & Knoblich, J. A. The Par complex and integrins direct asymmetric cell division in adult intestinal stem cells. *Cell Stem Cell* **11,** 529–540 (2012).
21. Sheng, X. R. & Matunis, E. Live imaging of the *Drosophila* spermatogonial stem cell niche reveals novel mechanisms regulating germline stem cell output. *Development* **138,** 3367–3376 (2011).
22. Zhu, Y., Huang, Y.-F., Kek, C. & Bulavin, D. V. Apoptosis differently affects lineage tracing of Lgr5 and Bmi1 intestinal stem cell populations. *Cell Stem Cell* **12,** 298–303 (2013).
23. Nakagawa, T., Sharma, M., Nabeshima, Y.-i., Braun, R. E. & Yoshida, S. Functional hierarchy and reversibility within the murine spermatogenic stem cell compartment. *Science* **328,** 62–67 (2010).
24. Klein, A. M., Nakagawa, T., Ichikawa, R., Yoshida, S. & Simons, B. D. Mouse germ line stem cells undergo rapid and stochastic turnover. *Cell Stem Cell* **7,** 214–224 (2010).
25. Rompolas, P. *et al.* Live imaging of stem cell and progeny behaviour in physiological hair-follicle regeneration. *Nature* **487,** 496–499 (2012).
26. Ritsma, L. *et al.* Surgical implantation of an abdominal imaging window for intravital microscopy. *Nature Protocols* **8,** 583–594 (2013).
27. Ritsma, L. *et al.* Intravital microscopy through an abdominal imaging window reveals a pre-micrometastasis stage during liver metastasis. *Sci. Transl. Med.* **4,** 158ra145 (2012).
28. Tian, H. *et al.* A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. *Nature* **478,** 255–259 (2011).
29. Rompolas, P., Mesa, K. R. & Greco, V. Spatial organization within a niche as a determinant of stem-cell fate. *Nature* **502,** 513–518 (2013).
30. Kozar, S. *et al.* Continuous clonal labeling reveals small numbers of functional stem cells in intestinal crypts and adenomas. *Cell Stem Cell* **13,** 626–633 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors would like to thank A. de Graaff from the Hubrecht Imaging Center for imaging support, all members of the van Rheenen group for useful discussions and the Hubrecht Institute animal caretakers for animal support. This work was supported by a Vidi fellowship (91710330; J.v.R.) and equipment grants (175.010.2007.00 and 834.11.002; J.v.R.) from the Dutch Organization of Scientific Research (NWO), a grant from the Dutch Cancer Society (KWF; HUBR 2009-4621; J.v.R.), a grant from the Association for International Cancer Research (AICR; 13-0297; J.v.R.), and the Wellcome Trust (grant number 098357/Z/12/Z; B.D.S.).

Author Contributions J.v.R. and L.R. conceived the study. L.R. optimized the surgical and imaging procedure. L.R., S.I.J.E., A.Z. and H.J.S. performed imaging experiments. L.R., H.J.S., B.D.S. and S.I.J.E. performed analyses. F.J.d.S. provided the Lgr5$^{DTR:eGFP}$ mice and B.D.S. did all biophysical modelling. L.R. and S.I.J.E. made the figures. J.v.R. and H.C. supervised the study. All authors discussed results and participated in preparation of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.v.R. (j.vanrheenen@hubrecht.eu) or H.C. (h.clevers@hubrecht.eu).

METHODS

Mice. All experiments were carried out in accordance with the guidelines of the Animal Welfare Committee of the Royal Netherlands Academy of Arts and Sciences. To obtain $Lgr5^{eGFP\text{-}Ires\text{-}CreERT2}$/R26R-Confetti mice, R26R-Confetti[7] mice were crossed with $Lgr5^{eGFP\text{-}Ires\text{-}CreERT2}$ mice[1]. Random double heterozygous male mice between 10 and 22 weeks old were used for experiments. Three days before imaging, mice were injected with 2.5–5 mg tamoxifen (single injection; Sigma Aldrich) to induce activation of Cre recombinase, which in turn induced expression of one of the Confetti colours (membranous CFP, cytoplasmic YFP and RFP). Nuclear GFP was also activated, but that subset of Confetti-labelled cells was not followed. For the targeted ablation studies, four male $Lgr5^{DTR:eGFP}$ mice with an AIW received $50\,\mu g\,kg^{-1}$ DT through intraperitoneal injections. Depletion of $Lgr5^+$ cells was confirmed by intravital imaging. Mice in which $Lgr5^+$ cells were not completely depleted after 24 h received a second DT injection. Mice were housed under standard laboratory conditions and received food and water *ad libitum*.

AIW surgery. The AIW surgery was performed as described previously[26]. In short, all surgical procedures were performed under 2% isoflurane (v/v) inhalation anaesthesia. Before surgery, buprenorphine (3 µg per mouse; Temgesic, BD Pharmaceutical Systems) was administered intramuscularly. The left lateral flank of the mice was shaved and the skin was disinfected with 70% (v/v) ethanol. Next, a left lateral flank incision was made through the skin and abdominal wall and a purse string suture was placed along the wound edge. A disinfected AIW (>1 h in 70% (v/v) ethanol) was placed glass-side down next to the mice and the ileum was placed on top. 3 M Vetbond Tissue Adhesive (*n*-butyl cyanoacrylate) was used to fix the ileum to the cover glass of the AIW and CyGel (BioStatus Limited) was added to diminish peristaltic movement. After 5 min the AIW was inverted and placed in the mouse, with the skin and abdominal wall placed inside the AIW groove. Then the sutures were tightened to stably secure the window into the animal. After surgery the mice were provided with food and water *ad libitum*. Furthermore, mice were closely monitored once a day before imaging for behaviour, reactivity, appearance and defecation.

Equipment and settings. Intravital imaging was performed on an inverted Leica TCS SP5 AOBS two-photon microscope with a chameleon Ti:Sapphire pumped optical parametric oscillator (Coherent) equipped with a ×25 (HCX IRAPO NA0.95 WD 2.5 mm) water objective and four non-descanned detectors (NDDs). The NDDs collect the following wavelengths: NDD1, <455 nm; NDD2, 455–490 nm; NDD3, 500–550 nm; NDD4, 560–650 nm. Sequential scanning was performed, exiting the tissue with 860 and 960 nm wavelengths. The Confetti colours were detected as follows: 860 nm, NDD2 (CFP and eGFP); 960 nm, NDD3 (eGFP and YFP), NDD4 (YFP and RFP). Second harmonic generation (SHG) signal was generated by 960 nm excitation at collagen I and detected in NDD2. Scanning was performed in a bidirectional mode at 700 Hz and 12 bit, with a zoom of 1.7, 512 × 512 pixels. z-stacks with 2.5 µm z-steps of typically 70–80 images were acquired. Re-identification of the same crypts over multiple days was accomplished by storing the *xy* coordinates of the imaged regions using the 'multiple position' function in the LAS-AF software and using the vasculature and the typical (Confetti) $Lgr5^+$ crypt pattern as visual landmarks.

Multi-day intestinal stem-cell imaging. After placing the AIW, mice were kept under anaesthesia and placed face-down in a custom-designed imaging box in which isoflurane (1% (v/v)) was administered through a facemask as described previously[26]. For the multi-day imaging sessions (Figs 1c, d, 2, 3b, 4 and Extended Data Figs 1, 2, 4, 5, 7), mice were imaged once a day for a maximum of 3 h, during which the climate chamber surrounding the microscope was kept at 32 °C. After the imaging session the mice were allowed to wake up to maintain their body temperature. After imaging, acquired z-stacks were corrected for *z* and *xy* shifts using a custom-designed Visual Basic software program and further processed and analysed using basic functions in ImageJ software (linear contrasting, blurring, median filtering).

Short-term intestinal stem-cell imaging. Mice were anaesthetized using isoflurane (2% (v/v)). The left lateral flank was shaved and the skin was disinfected using 70% (v/v) ethanol. Next, a left lateral flank incision was made through the skin and abdominal wall and the ileum was extracorporated using in PBS-drowned cotton swabs. The ileum was placed on a custom-designed inset containing a coverslip fitting the custom-designed imaging box. The ileum was secured to the coverslip using Vetbond and CyGel. The mouse was placed on top of the intestine and PBS-drowned sterile cotton gauzes were placed next to the animal to prevent dehydration. ParafilmM (Sigma Aldrich) was used to cover the mouse and a subcutaneous infusion system was used to provide 100 µl of sterile PBS per hour. The inset was placed within the custom-designed imaging box in which isoflurane (1% (v/v)) was administered through a facemask as described earlier. The temperature of the mouse was monitored during imaging using a rectal probe and was kept between 36 and 37 °C by adjusting the temperature of the surrounding climate chamber. Imaging was performed every 2 h for 14 h. z-stacks with a z-step of 2.5 µm of 12 regions with on average 6 crypts were made. Acquired z-stacks were analysed using ImageJ plugins (TurboReg, 3D visualization, 3D viewer).

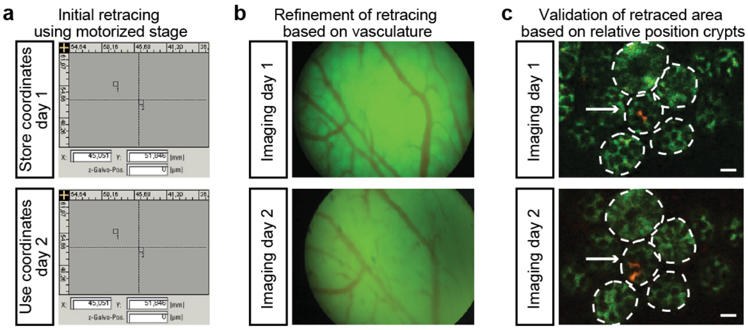
Intravital imaging of $Lgr5^+$-cell-depleted mice. Mice with an AIW received $50\,\mu g\,kg^{-1}$ DT through intraperitoneal injections. Depletion of $Lgr5^+$ cells was confirmed by intravital imaging. Mice in which $Lgr5^+$ cells were not completely depleted after 24 h received a second DT injection. Only mice in which full depletion was confirmed by intravital imaging 24 h after the last DT injection were analysed. The number of $Lgr5^+$–GFP cells within the stem-cell niche border and centre was determined as described in Fig. 1b.

Real-time lineage tracing of clonal competition. The data from the lineage tracing were collected at random, and all clones that were imaged with a 3-day interval were included. The strength of a Confetti-labelled $Lgr5^+$ CBC cell to produce offspring was expressed as the fold increase in Confetti-labelled $Lgr5^+$ CBC cell number 3 days after the first imaging session. A Mann–Whitney U test was performed because the data were not normally distributed.

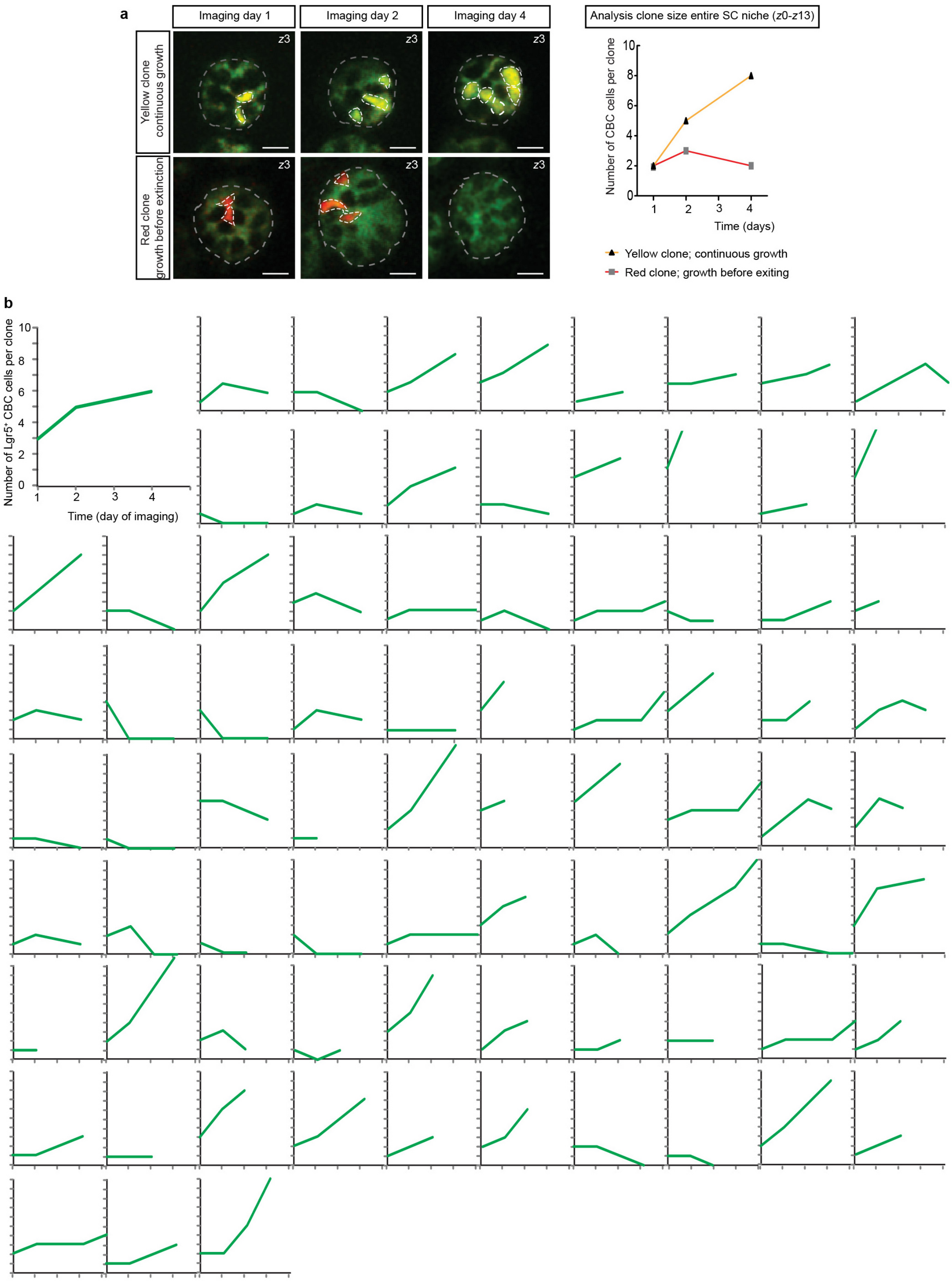
Quantitative data analysis of multi-day lineage tracing. Lineage tracing was performed for 80 clones in 80 crypts from 4 mice. No sample size estimate was calculated before the study was executed. Only data from mice from which high enough quality images were acquired were included in the study. The number of Confetti-labelled cells per crypt position (centre (rows 0 to +2), border (rows +3 to +4), TA (rows >4)) was scored. From the 80 lineages, we obtained 33 sublineages originating from the central region (Fig. 4a), and 47 sublineages from the border (Fig. 4b).

Immune cell analysis on intestinal tissue. Six E-Cadherin-CFP/$Lgr5^{eGFP\text{-}Ires\text{-}CreERT2}$ mice, 22 weeks of age, were randomly divided into two groups: a control and a window group to test potential side effects of AIWs. AIWs were implanted on top of the small intestines of mice from the window group, whereas mice from the control group did not undergo surgery. After 24 h, all mice were killed and the small intestines were harvested. Note that in the window group the part of the small intestine that was located directly behind the window was harvested. The small intestines were fixed for 1 day in fixation mix (1% paraformaldehyde, 0.2% $NaIO_4$, 61 mM Na_2HPO_4, 75 mM L-Lysine and 14 mM NaH_2PO_4 in H_2O). After fixation, the tissues were placed for 6 h into 30% sucrose after which the tissues were snap-frozen using Tissue Freezing Medium (Leice Microsystems Nussloch GmbH). Sixteen-micrometre sections were cut using a Leica CM3050 cryotome. A standard immunohistochemistry protocol was used to stain the sections with CD45 antibodies (BD Pharmingen, 553078, Clone 30-F11) and random areas were imaged. For analysis, ten areas within the imaged regions were selected and analysed in a blinded manner. The number of CD45-positive cells within a region was counted manually and an averaged number for each mouse was calculated. Next, the average of the three mice per group was calculated. A Mann–Whitney U test was performed because the sample was not distributed normally, and no significant differences were found. The variance between the groups was tested with an F-test, and was not different.

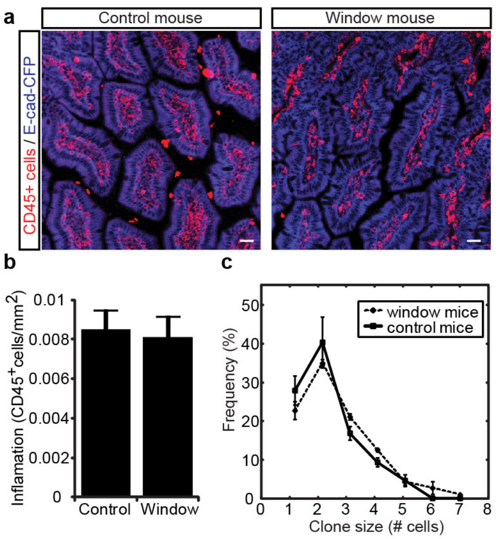
Clone frequency window versus control mice. Eight $Lgr5^{eGFP\text{-}Ires\text{-}CreERT2}$ mice, 22 weeks of age, were divided into two groups: a control and a window group. All mice received 5 mg tamoxifen by intraperitoneal injection. Three days later, AIWs were implanted on top of the small intestines of mice from the window group, whereas mice from the control group did not undergo surgery at this point. Two days after the surgery (5 days after tamoxifen injection) all mice were imaged. In the control group the intestine was exteriorized before imaging (as described earlier). In the window group the mice were imaged through the AIW. Several random areas were imaged. All recorded clones were used for analysis. For a single clone the number of cells within the stem-cell compartment was determined and a frequency distribution was made for the two groups.



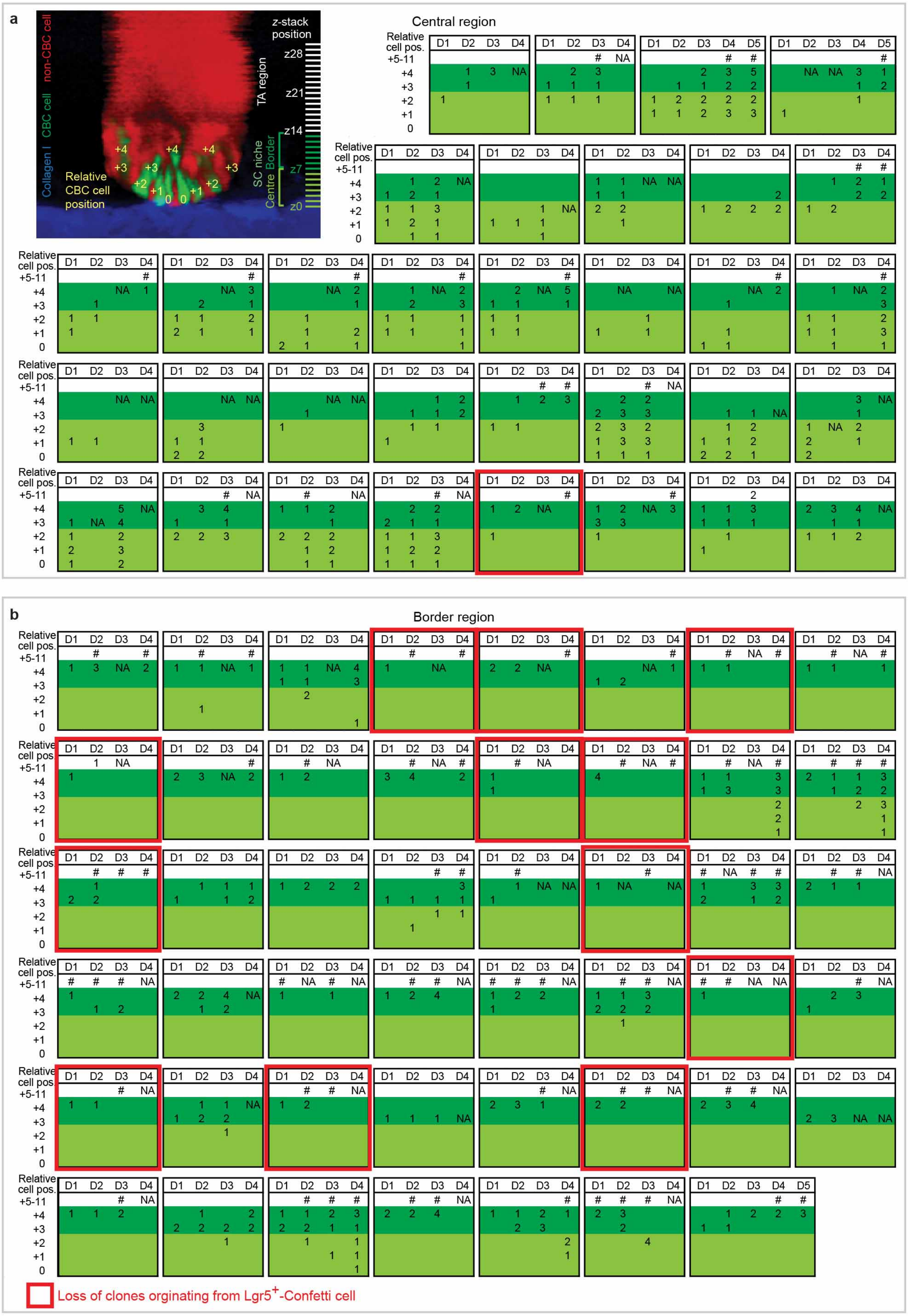
Extended Data Figure 1 | Retracing of the intravital imaging fields. **a**, The coordinates of the imaging fields within the imaging window (which always has a fixed position within the stage) were stored. By applying these stored coordinates in the subsequent imaging sessions, we recovered the same positions. **b**, The vasculature was used to refine the retracing. **c**, The retraced imaging areas were validated based on the relative position of the coloured crypts. Scale bars, 20 μm.



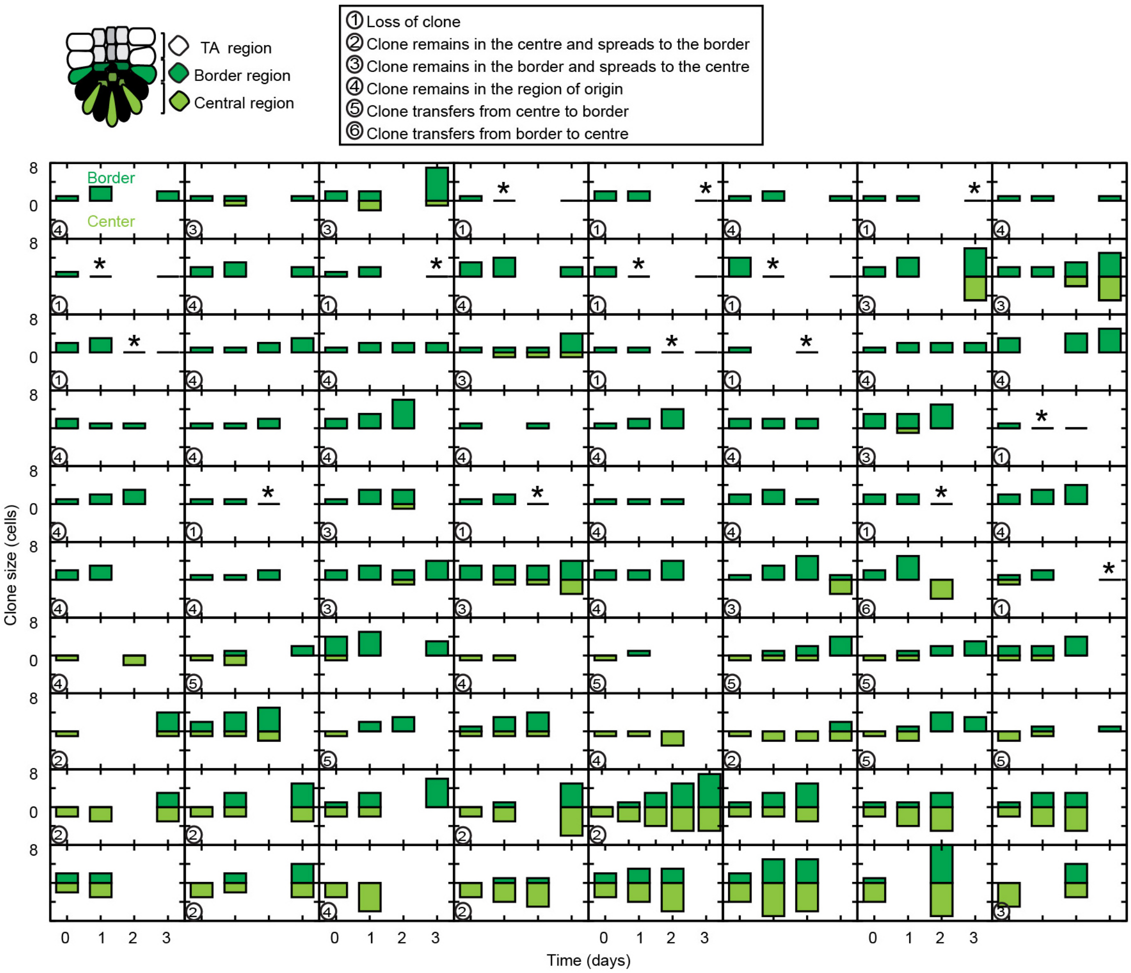
Extended Data Figure 2 | Real-time lineage tracing of individual clones. **a**, Expression of a Confetti colour was induced in Lgr5$^+$ CBC cells, and their progeny was followed over time. Two examples of Confetti-labelled Lgr5$^+$ CBC cells expanding over time are shown to illustrate clonal competition at indicated times. The top images show the continuous expansion of a YFP-expressing Lgr5$^+$ CBC cell (yellow line) in the central region of the crypt. The bottom images show the size of a clone derived from an RFP-expressing Lgr5$^+$ CBC cell in the central region of the crypt, which first increases and then declines (red line). The Confetti-labelled cells are outlined by a white dashed line and the crypts by a grey dashed line. Scale bars, 20 μm. Right, graph shows the increase in the number of YFP- or RFP-expressing Lgr5$^+$ CBC cells in the entire stem-cell compartment of the crypts shown in the images on the left (yellow and red line, respectively). **b**, A total of 80 crypts were imaged in four mice. The total clone size (border and central) of Confetti$^+$ cells present in the stem-cell niche independent of Lgr5 expression (row 0 to +4, which translates to z0 to z13 of the z-stack) was analysed at the indicated times ($n = 4$ mice) for all 80 lineages. The first time point translates to several days after induction. Every graph represents one crypt.



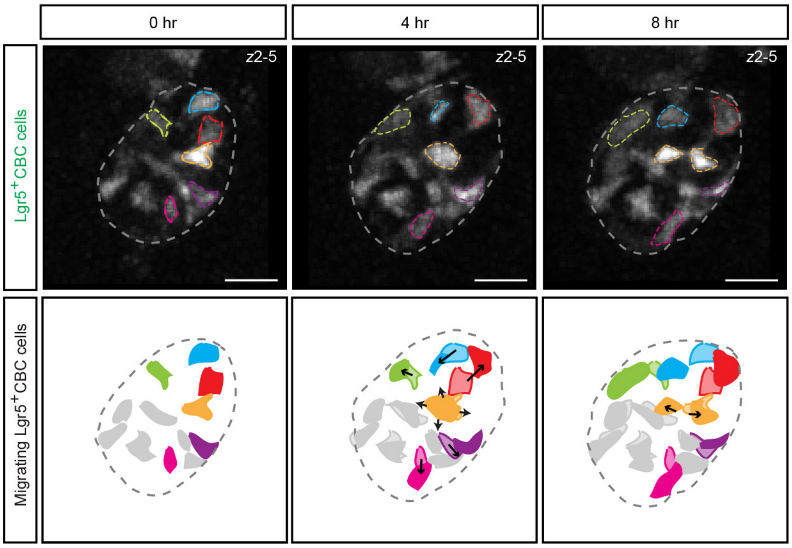
Extended Data Figure 3 | Clonal dynamics are unaffected by AIW surgery. **a**, Abdominal imaging windows (AIW) were surgically implanted into the abdominal wall of E-Cadherin-CFP/Lgr5$^{\text{eGFP-Ires-CreERT2}}$ mice. To detect CD45^{+} immune cells, tissue sections of the small intestine of these and control (no AIW) mice were stained with CD45 antibodies. Scale bar, 20 µm. **b**, Quantification of images in **a**. Regions of interest within the stromal area of the small intestines were measured and the number of CD45^{+} cells was counted within those regions. The bars show averages and s.e.m. ($n = 10$ regions per mouse performed in 3 mice per condition). **c**, Five days after tamoxifen injection, the clones of Lgr5$^{\text{eGFP-Ires-CreERT2}}$/R26R-Confetti mice were imaged intravitally. In control mice, the intestine was exteriorized before imaging and, in window mice, a window was placed 2 days before imaging. The frequency of clones with a certain size is plotted in the graph. The lines show the mean and s.e.m. ($n = 4$ mice per condition).



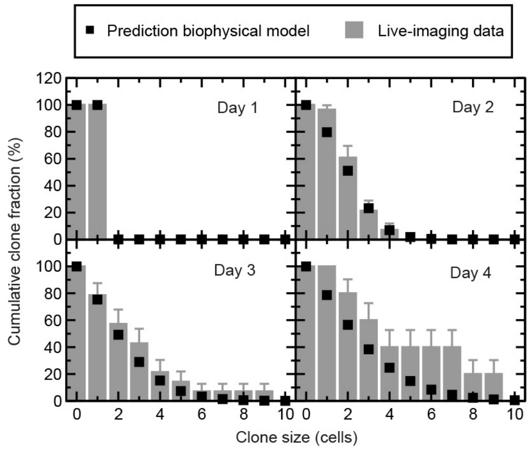
Extended Data Figure 4 | The spatial distribution of Confetti clone expansion within the stem-cell niche. **a**, **b**, Spatial distribution of confetti clones, subdivided into clones starting in the central (**a**) or border (**b**) region ($n = 4$ mice). Top left panel (**a**) shows the relative Confetti$^+$ cell position within the stem-cell niche, where row 0 to +4 translates to z0 to z13 of the z-stack. Every table shows the number of Confetti-labelled cells that are present at the different positions in the crypt, independent of Lgr5 expression. The position within the central and border region of the stem-cell niche and TA compartment are colour coded. The hash symbol indicates the presence of Confetti-labelled cells in the TA compartment. NA indicates time points for which data was not available owing to crypts that were not retraceable at that specific time point.



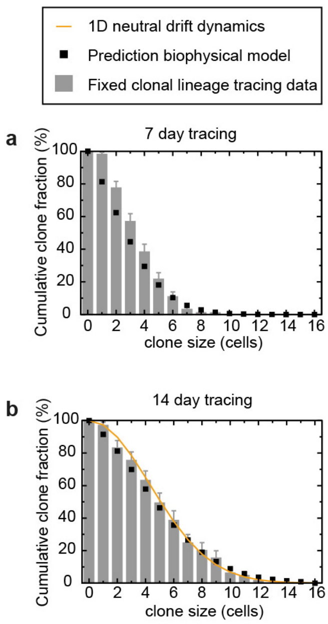
Extended Data Figure 5 | Average clone size and survival probability. Cartoon shows the division of the intestinal stem-cell niche into a central and border region. The central region contains rows 0 to +2, which translates to z0 to z6 of the z-stack, and the border region contains rows +3 and +4, which translates to z7 to z13 of the z-stack (see Fig. 1b). Confetti expression was induced in Lgr5^{+} CBC cells. The panels show, at the indicated time, the spatial organization of the 80 lineages of progeny of the Confetti-labelled CBC cells (same clones as in Extended Data Figs 2, 4) ($n = 4$ mice). For each lineage, we recorded the number of cells per clone in the central and border regions of the stem-cell niche at the indicated times. The first time point translates to several days after induction. The asterisks indicate clones in which all progeny were lost from the niche region.



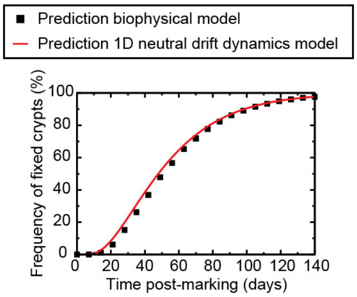
Extended Data Figure 6 | Lgr5^{+} CBC cells move and can become expelled from the stem-cell niche. Top, maximum projection images (z2 to z5) of a time series of a crypt. In the lower cartoons, the Lgr5^{+} CBC cells are highlighted. The moving cells are indicated with a different colour. The location of the cells at earlier time points are shown by a lighter colour and arrows indicate the direction of movement. Scale bars, 20 μm.



Extended Data Figure 7 | Cumulative clone size distributions derived from two border mother cells. Cumulative clone size distributions of clones derived from two neighbouring cells in the border region. In each case, the size of the clone is defined by the total number of constituent cells in the central and border regions (rows 0 to +4), independent of GFP expression. Note that the cumulative clone size distribution records the fraction of clones that have a size larger than the given value. The bars represent measurements from individual lineages reconstructed from live imaging at days 1 (**a**, $n = 13$; **b**, $n = 35$; **c**, $n = 35$), 2 (**a**, $n = 10$; **b**, $n = 25$; **c**, $n = 28$), 3 (**a**, $n = 7$; **b**, $n = 14$; **c**, $n = 14$) and 4 (**a**, $n = 6$; **b**, $n = 8$; **c**, $n = 6$). Error bars denote s.e.m. The points represent the predictions of the biophysical model using the same parameters as those inferred from the short-term live-imaging assay (for details, see Fig. 3a and Supplementary Notes).



Extended Data Figure 8 | Longer-term clonal evolution and the approach to scaling behaviour. **a**, **b**, Cumulative clone size distribution at 7 days (**a**) and 14 days (**b**) after induction, showing the percentage of clones that have a constituent number of $Lgr5^{+}$ CBC cells larger than the given value. For example, the data point at around four cells (40%) in **a** shows that, at 7 days post-induction, some 40% of clones have a size larger than four $Lgr5^{+}$ cells, and so on. The bars reproduce the findings of a fixed clonal assay using the same Lgr5-Confetti mouse construct used in the present study and previously described[7]. Error bars denote s.e.m. Points represent the predictions of the biophysical model using the same parameters as those inferred from the short-term live-imaging assay (see Fig. 3 and Supplementary Notes). To account for the prolonged activity of Cre recombinase in the static clonal assay, in both cases we introduced a 1-day time delay. At the 14-day time point, the measured clone size distribution and model prediction are beginning to converge onto the universal scaling behaviour characteristic of a strictly one-dimensional neutral drift dynamics (line) (for details, see Supplementary Notes).



Extended Data Figure 9 | Long-term drift towards monoclonality of labelled crypts. Predicted frequency of monoclonal crypts over time expressed as a percentage of surviving clones after pulse labelling of stem cells at clonal density. Points show the predictions of the biophysical model defined in the main text and Supplementary Notes using the same parameters as those inferred from the short-term live-imaging assay (see Fig. 3 and Supplementary Notes) after the representative marking of stem cells at the crypt base, and the line shows the predictions of the strictly one-dimensional neutral drift dynamics model introduced previously[8], with a stem-cell loss replacement rate of 0.24 per day and a total of 8 stem cells. The convergence of these two model predictions at longer times shows that, first, the behaviour of the quasi-one-dimensional model approaches that of the strictly one-dimensional model at longer times and, second, that the effective loss/replacement rate and stem-cell number in the new model is essentially fixed by the rate $P_{cc}\lambda = 0.24$ per day and the eight stem cells that occupy the central region. Significantly, these parameters translate to the ratio $P_{cc}\lambda/N_{stem}^2 = 0.026$ per week, very close to the figure of 0.025 per week obtained from a fit of the measured monoclonal crypt fraction to the one-dimensional neutral drift dynamics model in ref. 8. (For the labelling protocol and the experimental data points, we refer to the original reference.)

doi:10.1038/nature12979

IL-35-producing B cells are critical regulators of immunity during autoimmune and infectious diseases

Ping Shen[1]*, Toralf Roch[1]†*, Vicky Lampropoulou[1], Richard A. O'Connor[2], Ulrik Stervbo[1], Ellen Hilgenberg[1], Stefanie Ries[1], Van Duc Dang[1], Yarúa Jaimes[1], Capucine Daridon[1,3], Rui Li[4], Luc Jouneau[5], Pierre Boudinot[5], Siska Wilantri[1], Imme Sakwa[1], Yusei Miyazaki[4], Melanie D. Leech[2], Rhoanne C. McPherson[2], Stefan Wirtz[6], Markus Neurath[6], Kai Hoehlig[1], Edgar Meinl[7], Andreas Grützkau[1], Joachim R. Grün[1], Katharina Horn[1], Anja A. Kühl[8], Thomas Dörner[1,3], Amit Bar-Or[4], Stefan H. E. Kaufmann[9], Stephen M. Anderton[2] & Simon Fillatreau[1]

B lymphocytes have critical roles as positive and negative regulators of immunity. Their inhibitory function has been associated primarily with interleukin 10 (IL-10) because B-cell-derived IL-10 can protect against autoimmune disease and increase susceptibility to pathogens[1,2]. Here we identify IL-35-producing B cells as key players in the negative regulation of immunity. Mice in which only B cells did not express IL-35 lost their ability to recover from the T-cell-mediated demyelinating autoimmune disease experimental autoimmune encephalomyelitis (EAE). In contrast, these mice displayed a markedly improved resistance to infection with the intracellular bacterial pathogen *Salmonella enterica* serovar Typhimurium as shown by their superior containment of the bacterial growth and their prolonged survival after primary infection, and upon secondary challenge, compared to control mice. The increased immunity found in mice lacking IL-35 production by B cells was associated with a higher activation of macrophages and inflammatory T cells, as well as an increased function of B cells as antigen-presenting cells (APCs). During *Salmonella* infection, IL-35- and IL-10-producing B cells corresponded to two largely distinct sets of surface-IgM$^+$ CD138^{hi}TACI$^+$CXCR4$^+$CD1d^{int}Tim1^{int} plasma cells expressing the transcription factor Blimp1 (also known as Prdm1). During EAE, CD138$^+$ plasma cells were also the main source of B-cell-derived IL-35 and IL-10. Collectively, our data show the importance of IL-35-producing B cells in regulation of immunity and highlight IL-35 production by B cells as a potential therapeutic target for autoimmune and infectious diseases. This study reveals the central role of activated B cells, particularly plasma cells, and their production of cytokines in the regulation of immune responses in health and disease.

The inhibitory activities of B cells involve their production of IL-10, which in mice can protect from autoimmunity but impair resistance to infection[3–6]. Such suppressive function could be relevant to human diseases. A defect in IL-10 secretion by B cells was observed in patients with multiple sclerosis and type 1 diabetes[7,8]. Furthermore, B-cell depletion therapy had deleterious effects in some patients with multiple sclerosis or ulcerative colitis[9,10], and led to ulcerative colitis or psoriasis in patients with Graves' disease or rheumatoid arthritis, respectively[11,12]. These effects were probably not all owing to a loss of IL-10-producing B cells. Mouse B cells could inhibit immunity independently of IL-10 (refs 13, 14). However, no mediator to account for this has been characterized. There is an urgent need to identify additional factors mediating regulatory functions of B cells.

B cells require activation to exert suppressive activity and Toll-like receptors (TLR) are critical in this process. Mice with deficiencies in both TLR2 and TLR4 restricted to B cells developed an exacerbated EAE after immunization with the encephalitogenic peptide from myelin oligodendrocyte glycoprotein (MOG$_{35–55}$)[15]. Using mice with single deficiencies in these TLR restricted to B cells (B$^{Tlr2^{-/-}}$ and B$^{Tlr4^{-/-}}$ mice, respectively), we found that TLR4 was the most critical for B-cell-mediated suppression in EAE (Fig. 1a and Extended Data Fig. 1a). With our previous studies[3], these results establish TLR4 and CD40 as receptors essential for the regulatory function of B cells in EAE. CD40 also contributes to the protective roles of B cells in ulcerative colitis and arthritis[4,5].

IL-10 production by B cells is required for recovery from EAE[3]. Naive B cells produced IL-10 after TLR4 engagement, but not upon co-stimulation by TLR4 plus CD40 (Fig. 1b and Extended Data Fig. 1b). To identify additional suppressive factor(s) produced upon TLR4 plus CD40 stimulation, we performed Affymetrix array analyses on (1) naive B cells, (2) B cells activated by TLR4 and (3) B cells activated by TLR4 plus CD40, and focused on genes coding for secreted molecules (Extended Data Fig. 1c). Among the genes differentially expressed, of interest was Epstein-Barr virus-induced gene 3 (*Ebi3*), a member of the IL-12 cytokine family that can dimerize with p28 or p35 to generate IL-27 or IL-35, respectively, which both have suppressive functions[16–19]. B cells did not express *p40* (also known as *Il12b*) messenger RNA, highlighting its cell-type specific expression pattern[20], but constitutively transcribed *p35* (also known as *Il12a*) (Extended Data Fig. 1d). In fact, B cells were the main source of *p35* mRNA in secondary lymphoid tissues (Fig. 1d). B cells upregulated expression of *p35* and *Ebi3* mRNA, as well as EBI3 protein upon activation by TLR4 plus CD40, which was further increased upon B-cell receptor for antigen (BCR) engagement, suggesting they could secrete IL-35 (Fig. 1e and Extended Data Fig. 1). This was confirmed by co-immunoprecipitation using supernatants from TLR4 plus CD40-activated B cells (Fig. 1f). We conclude that B cells can secrete IL-35 after activation by TLR4 plus CD40.

To evaluate the role of IL-35 expression by B cells during EAE, we used mice with a B-cell-restricted deficiency in p35 (B$^{p35^{-/-}}$), or EBI3 (B$^{Ebi3^{-/-}}$), or control mice with wild-type B cells (B^{WT}). We also used mice in which B cells could not express p40 (B$^{p40^{-/-}}$), or p28 (B$^{p28^{-/-}}$), because p35 can dimerize with p40 to form IL-12, and EBI3 can associate with p28 to form IL-27. B cells secreted p28 after activation (Extended Data Fig. 1j). B$^{p35^{-/-}}$ and B$^{Ebi3^{-/-}}$ mice developed exacerbated EAE, whereas B$^{p40^{-/-}}$ and B$^{p28^{-/-}}$ mice had disease courses similar

[1]Deutsches Rheuma-Forschungszentrum, a Leibniz Institute, Charitéplatz 1, 10117 Berlin, Germany. [2]University of Edinburgh, Centre for Inflammation Research and Centre for Multiple Sclerosis Research, Queen's Medical Research Institute, Edinburgh EH16 4TJ, UK. [3]Charité Universitätsmedizin Berlin, CC12, Department of Medicine/Rheumatology and Clinical Immunology, 10117 Berlin, Germany. [4]Neuroimmunology Unit, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec H3A2B4, Canada. [5]Virologie et Immunologie Moléculaires, INRA, 78352 Jouy-en-Josas, France. [6]Medical Clinic 1, Kussmaul Campus for Medical Research, University of Erlangen-Nürnberg, 91054 Erlangen, Germany. [7]Institut für Klinische Neuroimmunologie Klinikum der Ludwig-Maximilians-Universität München, 81377 München, Germany. [8]Immunpathologie, Research Center ImmunoSciences, 12203 Berlin, Germany. [9]Max Planck Institute of Infection Biology, Department of Immunology, Charitéplatz 1, 10117 Berlin, Germany. †Present address: Institute of Biomaterial Science, Helmholtz-Zentrum Geesthacht, Centre for Materials and Coastal Research, Kantstraße 55, 14513 Teltow, Germany.
*These authors contributed equally to this work.

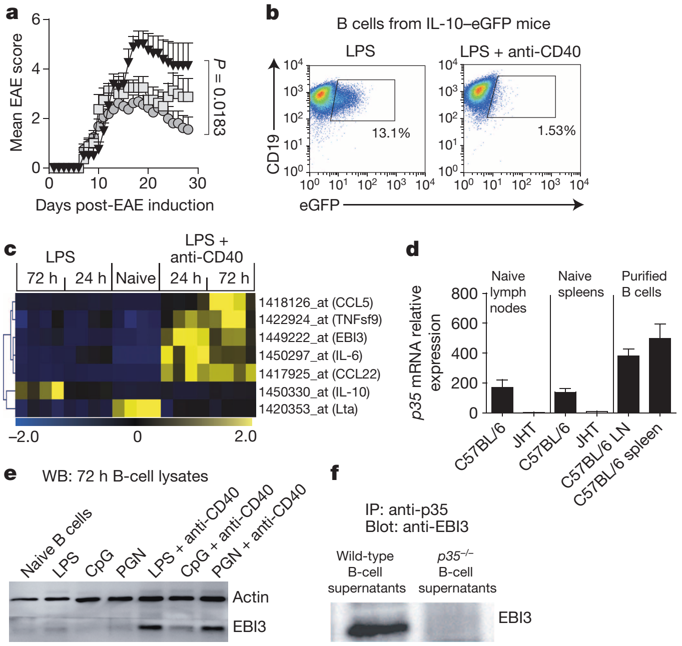


Figure 1 | B cells secrete IL-35 upon activation by TLR4 and CD40. **a**, EAE was induced in $B^{Tlr2^{-/-}}$ (grey squares, $n = 8$), $B^{Tlr4^{-/-}}$ (black triangles, $n = 8$) and B^{WT} mice (grey circles, $n = 16$). Data show clinical EAE scores from two independent experiments (mean ± s.e.m.). Cumulative disease scores were compared using two-tailed unpaired *t*-test. **b**, Splenic B cells from IL-10–eGFP (knock-in of eGFP at *Il10* locus) mice were stimulated for 72 h with LPS (1 μg ml^{-1}) or LPS (1 μg ml^{-1}) plus anti-CD40 (10 μg ml^{-1}), and eGFP expression was measured by flow cytometry. Plots show eGFP expression by live CD19^{+} cells. Results are representative of three independent experiments. **c**, Hierarchical cluster analysis of secreted factors differentially expressed between B cells activated with LPS or LPS plus anti-CD40 (Pearson correlation with average linkage). Affymetrix microarrays were performed in quadruplicates. Gene expression levels are shown for each array compared to its average value for all arrays, within a scale from twofold increase (yellow) to twofold decrease (blue). **d**, *p35* mRNA expression was quantified by RT–PCR in lymph nodes and spleen from naive C57BL/6 and B-cell-deficient JHT mice, as well as in B cells purified from lymph nodes and spleen of C57BL/6 mice. Data show compilation of three independent experiments (mean ± s.e.m.). **e**, Splenic B cells were activated as indicated for 72 h, and treated with GolgiStop for the last 4 h of culture. B-cell lysates were separated on SDS–PAGE gel and blotted with anti-EBI3 or anti-actin antibody. Data shows a representative result from three independent experiments. **f**, B cells from C57BL/6 or p35-deficient mice were activated for 72 h with LPS plus anti-CD40. Culture supernatants were subjected to immunoprecipitation with anti-p35 followed by western blot with anti-EBI3 antibody. Data shown are representative of two independent experiments.

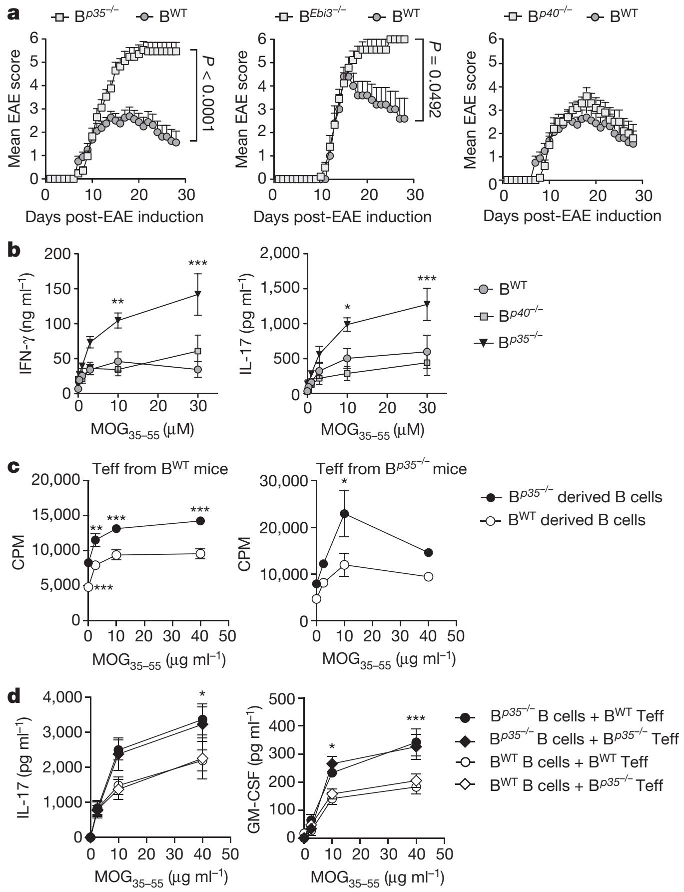


Figure 2 | IL-35 expression by B cells is required for recovery from EAE. **a**, EAE was induced in $B^{p35^{-/-}}$ (grey squares, $n = 17$) and B^{WT} mice (black circles, $n = 16$) (left panel); $B^{Ebi3^{-/-}}$ (grey squares, $n = 9$) and corresponding B^{WT} mice (black circles, $n = 5$) (middle panel); $B^{p40^{-/-}}$ (grey squares, $n = 10$) and B^{WT} mice (black circles, $n = 16$) (right panel) by immunization with $MOG_{35–55}$ peptide in Complete Freund's adjuvant. Data show clinical EAE scores from two independent experiments (mean ± s.e.m.). Cumulative disease scores were compared using a two-tailed unpaired *t*-test. **b**, Splenocytes were collected from mice on day 10 after EAE induction, and pooled before re-stimulation for 48 h with $MOG_{35–55}$ in increasing concentrations. Culture supernatants were analysed by ELISA to determine IFN-γ and IL-17 concentrations. Data show a representative result from two independent experiments. **c**, EAE was induced in $B^{p35^{-/-}}$ and corresponding B^{WT} mice by immunization with $MOG_{35–55}$ peptide in Complete Freund's adjuvant. B cells and CD4^{+}CD25^{-} T cells (Teff) were isolated from pooled draining lymph nodes and spleens on day 10 after immunization. Then 5×10^5 B cells from $B^{p35^{-/-}}$ or B^{WT} mice were cultured with 1×10^4 Teff cells from $B^{p35^{-/-}}$ or B^{WT} mice in the presence of $MOG_{35–55}$ in increasing concentrations, as indicated. Proliferation was assessed after 64 h by [^{3}H]thymidine incorporation. CPM, counts per minute. Data show representative results from two independent experiments. **d**, Supernatants from cultures as described in **c** were harvested after 48 h, and analysed by Bio-Plex to determine the concentrations of IL-17 and GM-CSF. Data shown (mean ± s.e.m.) are pooled from two independent experiments. **b–d**, Graphs show mean ± s.e.m.. Results were compared using a two-way ANOVA followed by a Bonferroni post-test (*$P < 0.05$, **$P < 0.01$; ***$P < 0.001$). Results of comparison are shown for $B^{p35^{-/-}}$ versus B^{WT} (**b**), and for B^{WT} B cells plus B^{WT} Teff versus $B^{p35^{-/-}}$ B cells plus B^{WT} Teff (**d**).

to B^{WT} controls (Fig. 2a and Extended Data Fig. 2). Therefore, B cells limited EAE pathogenesis through provision of IL-35. EAE pathogenesis involves T_H1 and T_H17 cells, which express IFN-γ and IL-17, respectively[21,22]. $B^{p35^{-/-}}$ mice displayed increased MOG-reactive IFN-γ and IL-17 production compared to control mice (Fig. 2b and Extended Data Fig. 2d). In contrast, $B^{p28^{-/-}}$ and $B^{p40^{-/-}}$ mice mounted normal T-cell responses (Fig. 2b and Extended Data Fig. 2). $B^{p35^{-/-}}$ mice had more CD4^{+} T cells and mononuclear phagocytes in the central nervous system than B^{WT} mice (Extended Data Fig. 2), suggesting that B-cell-derived IL-35 limited disease by reducing the accumulation of pathogenic cells in the target organ. These data demonstrate that B-cell-derived IL-35 is a critical regulator of T-cell-mediated autoimmunity.

The mechanisms underlying the suppressive activities of IL-35 remain poorly understood[23]. During EAE, the increased T-cell response observed in $B^{p35^{-/-}}$ mice was not due to a defect in CD4^{+}Foxp3^{+} T regulatory (Treg) cells (Extended Data Fig. 2), which are protective in this disease[24]. B cells from $B^{p35^{-/-}}$ mice expressed higher levels of activation markers (CD44, CD69) and molecules involved in antigen presentation to CD4^{+} T cells (MHC-II, CD80, CD86), compared to control B cells (Extended Data Fig. 3). Accordingly, they were more potent APC, stimulating higher proliferation and production of inflammatory cytokines (IL-17 and GM-CSF) by MOG-reactive CD4^{+} T cells than control B cells (Fig. 2c, d and Extended Data Fig. 3c). Preliminary studies showed that $B^{p35^{-/-}}$ mice were susceptible to EAE induced with human MOG, a

disease dependent on pathogenic functions of B cells[25]. These findings introduce IL-35 as a regulator of the APC function of B cells.

To test the role of B-cell-derived IL-35 in infection, we challenged $B^{p35-/-}$, $B^{Ebi3-/-}$ and $B^{p40-/-}$ mice with the Gram-negative bacterium *Salmonella enterica* serovar Typhimurium (hereafter referred to as *Salmonella*). This intracellular pathogen causes a disease in mice that resembles typhoid fever in humans, responsible for approximately 20 million cases and 600,000 deaths annually[26]. $B^{p35-/-}$ and $B^{Ebi3-/-}$ mice displayed longer survival than $B^{p40-/-}$ and B^{WT} mice after primary infection, and upon secondary challenge (Fig. 3a). This improved resistance correlated with better control of the bacterial burden in spleen and liver, increased accumulation of macrophages in these organs, and stronger inflammatory T-cell responses towards *Salmonella* (Fig. 3b, c and Extended Data Fig. 4). In contrast, B-cell-derived IL-35 had no effect on Treg frequencies, global frequencies of activated T cells or humoral immunity against *Salmonella* (Extended Data Figs 4 and 5a). $B^{p35-/-}$ mice also mounted normal antibody responses against a hapten-protein antigen (Extended Data Fig. 5b). Consistent with the role of CD40 in IL-35 production by B cells (Fig. 1), mice with a B-cell-restricted deficiency in CD40 displayed enhanced control of *Salmonella* infection (Extended Data Fig. 5c). These data demonstrate that B cells can inhibit anti-microbial immunity through production of IL-35.

B cells can also inhibit anti-*Salmonella* immunity through IL-10 (ref. 6). To identify IL-10- and IL-35-producing B cells, and clarify their relationship, we quantified *Ebi3* and *Il10* mRNA in $CD19^+CD138^-$ B cells and $CD138^{hi}$ plasma cells during *Salmonella* infection (Fig. 4a and Extended Data Fig. 6). *Ebi3* and *Il10* were exclusively induced in $CD138^{hi}$ cells (Fig. 4a). Single-cell PCR analyses indicated that distinct sets of $CD138^{hi}$ cells expressed the mRNA for *Il10* or for both IL-35 subunits *Ebi3* and *p35* (Fig. 4b). We therefore characterized further these plasma cells. They expressed uniform surface levels of IgM, CD80, CD86, MHC-II, CD40, CD69, CD44, CD43, TACI and CXCR4, as well as intermediate levels of CD1d (also known as Cd1d1) and Tim1 (also known as Havcr1) (Extended Data Fig. 6a, d), yet three subsets could be distinguished according to CD138 and CD22 levels: $CD138^{int}CD22^+$, $CD138^{hi}CD22^+$ and $CD138^{hi}CD22^-$ cells (Extended Data Fig. 7). These subsets differed by their capacity to produce antibodies, and expressed distinct amounts of mRNA for the transcription factors driving plasma cell development (Blimp1, IRF4), or maintaining B-cell identity (Pax5) (Extended Data Fig. 7), demonstrating that they corresponded to different stages of plasma cell development. The expression levels of *Il10* and *Ebi3* mRNA in these subsets correlated with their degree of maturity, and were highest in the most differentiated $CD138^{hi}CD22^-$ cells (Extended Data Fig. 7c). $CD138^{hi}CD22^-$ plasma cells were mostly located at the interface between red and white pulp in spleen, in clusters also containing T cells, $CD11b^+$ and $CD169^+$ myeloid cells (Extended Data Fig. 7d). Single-cell PCR analyses revealed that 6–10% of $CD138^{hi}$ cells expressed *Il10* mRNA, and a similar frequency made mRNA for both IL-35 subunits, whereas few cells co-expressed these three transcripts together (Fig. 4b). Nearly all $CD138^{hi}$ cells transcribing *Il10* or both IL-35 subunits co-expressed *Blimp1* (Fig. 4c), as expected for plasma cells. The less mature $CD138^{int}CD22^+$ population contained little *Il10* or *Ebi3* mRNA (Extended Data Fig. 7c), and only rare cells contained *Il10* mRNA (Fig. 4b), suggesting that expression of IL-10 and both IL-35 subunits were acquired during plasma cell maturation. These data indicate that distinct sets of plasma cells provide IL-10 and IL-35 during *Salmonella* infection.

Plasma cells from mice infected with *Salmonella* consistently secreted more IL-10 than B cells upon *ex vivo* stimulation (Fig. 4d and Extended Data Fig. 8). Around 70% of $CD138^{hi}$ plasma cells upregulated IL-10 expression after stimulation with phorbol 12-myristate 13-acetate plus ionomycin, a classical treatment for identifying cytokine-producing cells (Extended Data Fig. 8). $CD138^+$ plasma cells were also the major B-cell subtype expressing the proteins EBI3 and p35 (Fig. 4e). In contrast, $CD138^{hi}$ plasma cells did not secrete IL-6, a mediator of pro-inflammatory functions of B cells[27], unlike $CD19^+CD138^-$ B cells (Extended Data Fig. 8). This lack of IL-6 production, which may reflect a repressive effect of Blimp1 (ref. 28), distinguishes IL-10- and IL-35-expressing plasma cells from IL-10-producing $CD1d^{hi}$ B cells (Extended Data Fig. 8). $CD138^{hi}$ plasma cells therefore have a remarkable propensity to express anti-inflammatory cytokines but not IL-6 during *Salmonella*

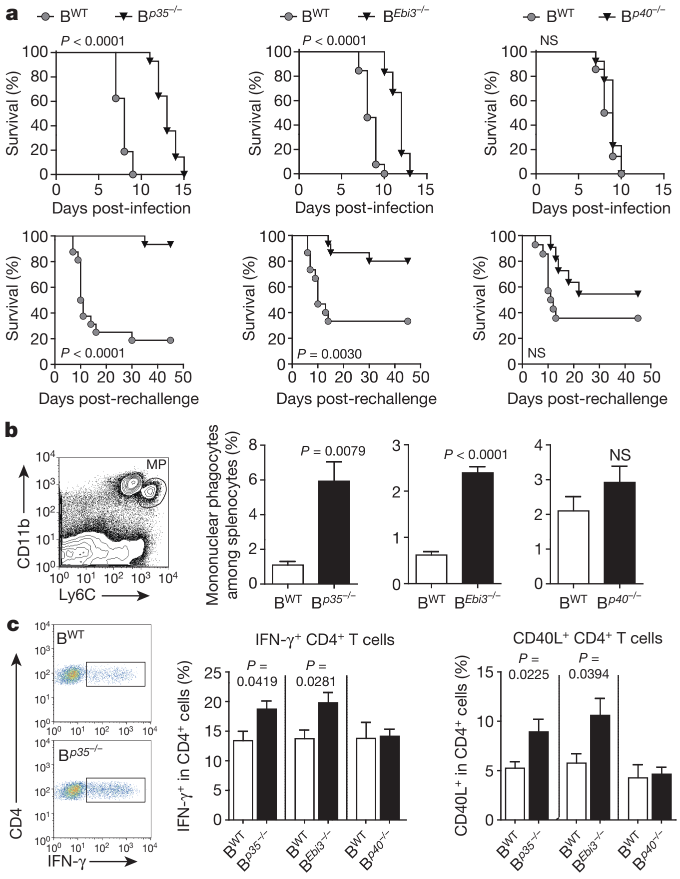


Figure 3 | B-cell-derived IL-35 enhances susceptibility to *Salmonella*. **a**, Top panel shows survival curves of $B^{p35-/-}$ ($n = 14$) and corresponding B^{WT} mice ($n = 16$), $B^{Ebi3-/-}$ ($n = 12$) and corresponding B^{WT} mice ($n = 13$), $B^{p40-/-}$ ($n = 13$) and corresponding B^{WT} mice ($n = 14$) after infection with virulent *Salmonella* (SL1344). Data pooled from two independent experiments. Survival curves were compared using Wilcoxon test. Bottom panel shows survival curves of $B^{p35-/-}$ ($n = 15$) and corresponding B^{WT} mice ($n = 16$), $B^{Ebi3-/-}$ ($n = 15$) and corresponding B^{WT} mice ($n = 15$), $B^{p40-/-}$ ($n = 11$) and corresponding B^{WT} mice ($n = 14$) vaccinated with attenuated *Salmonella* (SL7207) and 90 days later re-challenged with virulent *Salmonella* (SL1344). Data are pooled from two independent experiments. Survival curves were compared using Wilcoxon test. **b**, Representative FACS plot of mononuclear phagocytes (MP) gated as $CD11b^+Ly6C^{hi}$ cells among live B^{WT} splenocytes at day 6 post-infection (p.i). with SL1344 (left panel); frequencies of MP per spleen at day 6 p.i. in $B^{p35-/-}$, $B^{Ebi3-/-}$, and $B^{p40-/-}$ mice together with their corresponding B^{WT} controls (right panel). Numbers of mice: $B^{p35-/-}$ ($n = 6$) and B^{WT} ($n = 8$), $B^{Ebi3-/-}$ ($n = 7$) and B^{WT} ($n = 5$), $B^{p40-/-}$ ($n = 8$) and B^{WT} ($n = 8$). Data are pooled from two independent experiments. Graphs show mean ± s.e.m. **c**, Mice were infected with attenuated *Salmonella* (SL7207). After 21 days, bone marrow cells were stained for CD4 and CD40L or IFN-γ after 6 h re-stimulation with heat-killed *Salmonella*. Representative FACS plots of IFN-γ^+ cells among $CD4^+$ T cells from B^{WT} and $B^{p35-/-}$ mice (left panel). Frequencies of IFN-γ^+ and $CD40L^+$ cells among $CD4^+$ T cells, (middle and right panels, respectively). Data are pooled from two independent experiments with total number of mice: $B^{p35-/-}$ ($n = 12$) and corresponding B^{WT} ($n = 10$), $B^{Ebi3-/-}$ ($n = 8$) and corresponding B^{WT} ($n = 9$), $B^{p40-/-}$ ($n = 15$) and corresponding B^{WT} ($n = 9$). **b**, **c**, Data were analysed using two-tailed unpaired *t*-test with Welch's correction in case of unequal variance. *P* values >0.05 are considered not significant (NS).

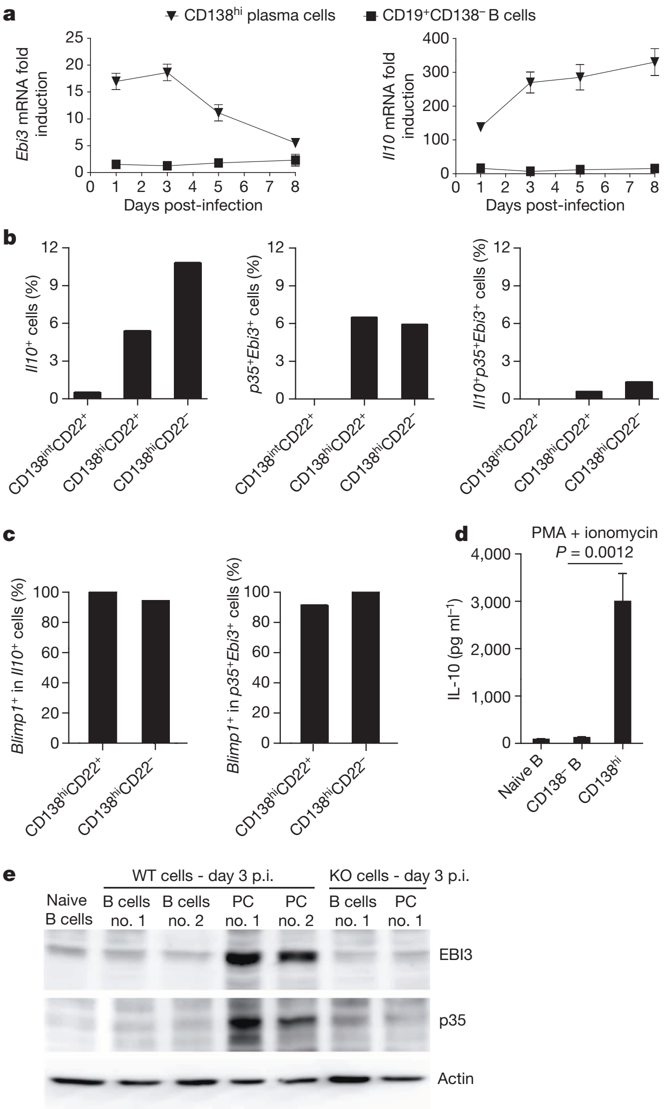


Figure 4 | IL-10 and IL-35 are expressed by CD138^{hi} plasma cells during *Salmonella* infection. **a**, Splenic plasma cells (PC) (CD138^{hi}) and B cells (CD19^{+}CD138^{-}) were isolated from C57BL/6 mice on days 0, 1, 3, 5, and 8 after infection with 10^7 c.f.u. attenuated *Salmonella* (SL7207). *Ebi3* and *Il10* mRNA expression was quantified by RT–PCR. Data show fold induction of *Ebi3* (left) and *Il10* (right) in PC and B cells compared to naive B cells. A compilation of five independent experiments is shown (mean ± s.e.m.). **b**, Single CD138^{int}CD22^{+}, CD138^{hi}CD22^{+}, and CD138^{hi}CD22^{-} cells were sorted by FACS from C57BL/6 mice on day 3 after infection with 10^7 c.f.u. attenuated *Salmonella* (SL7207). A total of 208 CD138^{int}CD22^{+}, 206 CD138^{hi}CD22^{+} and 189 CD138^{+}CD22^{-} cells expressed β-actin, and were included for analysis. Data show percentages of *Il10*$^{+}$ (left), *p35*$^{+}$ *Ebi3*$^{+}$ (middle), and *Il10*$^{+}$*p35*$^{+}$*Ebi3*$^{+}$ (right) cells among each subset. **c**, *Blimp1* mRNA expression in cells analysed in (**c**) was also detected by single-cell PCR. Data show percentages of *Blimp1*$^{+}$ cells among *Il10*$^{+}$ (left), and *p35*$^{+}$*Ebi3*$^{+}$ cells (right). **d**, CD138^{hi} PC and CD19^{+}CD138^{-} B cells were isolated from spleens of C57BL/6 mice on day 3 after infection with attenuated *Salmonella* (10^7 c.f.u.). Naive B splenic B cells were isolated from unchallenged C57BL/6 mice. Cells were activated for 24 h with PMA/ionomycin and IL-10 was determined in culture supernatants by Bio-Plex. Data shown is pooled from 5 independent experiments. Results were compared using a two-tailed unpaired *t*-test. **e**, Splenic CD19^{+}CD138^{-} B cells and CD138^{+} plasma cells were isolated from spleens of C57BL/6 (WT) and *p35*$^{-/-}$*Ebi3*$^{-/-}$*p40*$^{-/-}$ mice on day 3 p.i. with attenuated *Salmonella* (10^7 c.f.u.). Proteins were separated on SDS–PAGE gel and detected with anti-EBI3, anti-p35 or anti-actin antibodies. Data show results from two independent preparations for WT samples, and one preparation for *p35*$^{-/-}$*Ebi3*$^{-/-}$*p40*$^{-/-}$ B and plasma cells.

infection, emphasizing the regulatory potential of plasma cells compared to other B cell subsets.

Our study identifies IL-35-producing B cells as critical regulators of immunity. At a time window during *Salmonella* infection when both B-cell-derived IL-10 (ref. 6) and IL-35 exerted suppressive functions, plasma cells were the major B-cell type expressing these cytokines. This was also the case during EAE (Extended Data Fig. 9). During *Salmonella* infection, IL-10 and IL-35 were made by distinct sets of plasma cells, suggesting that these two suppressive axes can operate in parallel. In line with this, mice in which individual B cells could express either IL-10 or IL-35, but not both cytokines, displayed a normal EAE course (Extended Data Fig. 9). Accordingly, B and plasma cells could produce IL-10 without concomitant IL-35 secretion (Extended Data Fig. 9). Future studies shall assess whether 'regulatory plasma cells' producing IL-10 and IL-35 (but not IL-6) can provide novel opportunities for immune intervention.

METHODS SUMMARY

$B^{Tlr2-/-}$, $B^{Tlr4-/-}$, $B^{p35-/-}$, $B^{Ebi3-/-}$, $B^{p40-/-}$, $B^{p28-/-}$ and B^{WT} mice were generated using a previously described mixed bone marrow chimaera approach[3,6], as detailed in the Methods, and outlined in Extended Data Fig. 10. This protocol allowed normal reconstitution of the B-cell compartment in the resulting chimaera (Extended Data Fig. 10). EAE was induced in mice by immunization with MOG$_{35-55}$ emulsified in complete Freund's adjuvant (Sigma-Aldrich). The disease score and the MOG$_{35-55}$ specific T-cell responses (at days 10 and 28) were assessed as described previously[3]. Mouse B cells were purified and activated as described before[15]. Cytokine secretion was determined by Bio-Plex or ELISA. Whole transcriptome analyses were performed with RNA purified from naive and activated B cells, using GeneChip Mouse Genome 430 2.0 Arrays. For mRNA quantification, total RNA was isolated with TRIzol (AMS Biotechnology) and cDNA was prepared with the Reverse Transcription Kit (Promega). Quantitative RT–PCR was performed with LightCycler FastStart DNA Master SYBR Green I (Roche). Western blots to determine EBI3, p35 and actin expression were performed using anti-EBI3 (rabbit M75 polyclonal Ab), anti-p35 (rabbit EPR5736 polyclonal Ab), and anti-actin (rabbit I-19 polyclonal Ab), respectively. For immunoprecipitation, IL-35 was captured with anti-p35 antibody (clone C18.2; eBioscience) followed by precipitation using µMACS Protein G microbeads (Miltenyi Biotech). Infections with virulent and attenuated *Salmonella* (strains SL1344 and SL7207, respectively) were performed as published[6]. Flow cytometry analyses were performed as previously described[6]. Other protocols are described in the Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 January 2012; accepted 30 December 2013.
Published online 23 February 2014.

1. Fillatreau, S., Gray, D. & Anderton, S. M. Not always the bad guys: B cells as regulators of autoimmune pathology. *Nature Rev. Immunol.* **8,** 391–397 (2008).
2. Fillatreau, S. Novel regulatory functions for Toll-like receptor-activated B cells during intracellular bacterial infection. *Immunol. Rev.* **240,** 52–71 (2011).
3. Fillatreau, S., Sweenie, C. H., McGeachy, M. J., Gray, D. & Anderton, S. M. B cells regulate autoimmunity by provision of IL-10. *Nature Immunol.* **3,** 944–950 (2002).
4. Mauri, C., Gray, D., Mushtaq, N. & Londei, M. Prevention of arthritis by interleukin 10-producing B cells. *J. Exp. Med.* **197,** 489–501 (2003).
5. Mizoguchi, A., Mizoguchi, E., Takedatsu, H., Blumberg, R. S. & Bhan, A. K. Chronic intestinal inflammatory condition generates IL-10-producing regulatory B cell subset characterized by CD1d upregulation. *Immunity* **16,** 219–230 (2002).
6. Neves, P. *et al.* Signaling via the MyD88 adaptor protein in B cells suppresses protective immunity during *Salmonella typhimurium* infection. *Immunity* **33,** 777–790 (2010).
7. Duddy, M. *et al.* Distinct effector cytokine profiles of memory and naive human B cell subsets and implication in multiple sclerosis. *J. Immunol.* **178,** 6092–6099 (2007).
8. Jagannathan, M. *et al.* Toll-like receptors regulate B cell cytokine production in patients with diabetes. *Diabetologia* **53,** 1461–1471 (2010).
9. Benedetti, L. *et al.* Relapses after treatment with rituximab in a patient with multiple sclerosis and anti myelin-associated glycoprotein polyneuropathy. *Arch. Neurol.* **64,** 1531–1533 (2007).
10. Goetz, M., Atreya, R., Ghalibafian, M., Galle, P. R. & Neurath, M. F. Exacerbation of ulcerative colitis after rituximab salvage therapy. *Inflamm. Bowel Dis.* **13,** 1365–1368 (2007).
11. El Fassi, D., Nielsen, C. H., Kjeldsen, J., Clemmensen, O. & Hegedus, L. Ulcerative colitis following B lymphocyte depletion with rituximab in a patient with Graves' disease. *Gut* **57,** 714–715 (2008).

12. Dass, S., Vital, E. M. & Emery, P. Development of psoriasis after B cell depletion with rituximab. *Arthritis Rheum.* **56,** 2715–2718 (2007).
13. Wilson, M. S. *et al.* Helminth-induced CD19^{+}CD23^{hi} B cells modulate experimental allergic and autoimmune inflammation. *Eur. J. Immunol.* **40,** 1682–1696 (2010).
14. Su, Y., Zhang, A. H., Noben-Trauth, N. & Scott, D. W. B-cell gene therapy for tolerance induction: host but not donor B-cell derived IL-10 is necessary for tolerance. *Front. Microbiology* **2**, http://dx.doi.org/10.3389/fmicb.2011.00154 (2011).
15. Lampropoulou, V. *et al.* TLR-activated B cells suppress T cell-mediated autoimmunity. *J. Immunol.* **180,** 4763–4773 (2008).
16. Devergne, O., Birkenbach, M. & Kieff, E. Epstein-Barr virus-induced gene 3 and the p35 subunit of interleukin 12 form a novel heterodimeric hematopoietin. *Proc. Natl Acad. Sci. USA* **94,** 12041–12046 (1997).
17. Collison, L. W. *et al.* The inhibitory cytokine IL-35 contributes to regulatory T-cell function. *Nature* **450,** 566–569 (2007).
18. Niedbala, W. *et al.* IL-35 is a novel cytokine with therapeutic effects against collagen-induced arthritis through the expansion of regulatory T cells and suppression of Th17 cells. *Eur. J. Immunol.* **37,** 3021–3029 (2007).
19. Villarino, A. *et al.* The IL-27R (WSX-1) is required to suppress T cell hyperactivity during infection. *Immunity* **19,** 645–655 (2003).
20. Brentano, F. *et al.* Abundant expression of the interleukin (IL)23 subunit p19, but low levels of bioactive IL23 in the rheumatoid synovium: differential expression and Toll-like receptor-(TLR) dependent regulation of the IL23 subunits, p19 and p40, in rheumatoid arthritis. *Ann. Rheum. Dis.* **68,** 143–150 (2009).
21. Kuchroo, V. K. *et al.* Cytokines and adhesion molecules contribute to the ability of myelin proteolipid protein-specific T cell clones to mediate experimental allergic encephalomyelitis. *J. Immunol.* **151,** 4371–4382 (1993).
22. Park, H. *et al.* A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin 17. *Nature Immunol.* **6,** 1133–1141 (2005).
23. Bettini, M., Castellaw, A. H., Lennon, G. P., Burton, A. R. & Vignali, D. A. Prevention of autoimmune diabetes by ectopic pancreatic β-cell expression of interleukin-35. *Diabetes* **61,** 1519–1526 (2012).
24. McGeachy, M. J., Stephens, L. A. & Anderton, S. M. Natural recovery and protection from autoimmune encephalomyelitis: contribution of CD4^{+}CD25^{+} regulatory cells within the central nervous system. *J. Immunol.* **175,** 3025–3032 (2005).
25. Weber, M. S. *et al.* B-cell activation influences T-cell polarization and outcome of anti-CD20 B-cell depletion in central nervous system autoimmunity. *Ann. Neurol.* **68,** 369–383 (2010).
26. Mittrücker, H. W. & Kaufmann, S. H. Immune response to infection with *Salmonella typhimurium* in mice. *J. Leukoc. Biol.* **67,** 457–463 (2000).
27. Barr, T. A. *et al.* B cell depletion therapy ameliorates autoimmune disease through ablation of IL-6-producing B cells. *J. Exp. Med.* **209,** 1001–1010 (2012).
28. Chan, Y. H. *et al.* Absence of the transcriptional repressor Blimp-1 in hematopoietic lineages reveals its role in dendritic cell homeostatic development and function. *J. Immunol.* **183,** 7039–7046 (2009).

Acknowledgements We thank H. Schliemann, H. Ruebsamen, M. Spadaro and D. Jenne for assistance and support. We thank M. Loehning for providing IL-12 p40-deficient mice, and S. Akira for providing TLR2- and TLR4-deficient mice. We thank O. Neyrolles for help with some of the *in vivo* experiments. We thank E. Schott for help with the AST/ALT measurements. S.F. is supported by grants from the Deutsche Forschungsgemeinschaft (SFB-650, TRR-36, TRR-130, FI-1238/02), Hertie Stiftung, and an advanced grant from the Merieux Institute. C.D and T.D. are supported by the Deutsche Forschungsgemeinschaft (SFB-650, Do491/7-2, 8-2). P.B. and L.J. are supported by INRA. A.B.-O. is supported by a CIHR/MSSC New Emerging Team grant in Clinical Autoimmunity. Work in S.M.A.'s laboratory was supported by grants from the UK Medical Research Council and the Wellcome Trust. E.M. is supported by the Clinical Competence Network for Multiple Sclerosis and SFB-TR128.

Author Contributions P.S., T.R. and V.L. performed most of the experiments, the data analysis, and edited the manuscript. R.A.O., U.S., E.H., S.R., V.D.D., Y.J., C.D., R.L., L.J., P.B., S.W., I.S., Y.M., M.D.L., R.C.M., S.W., M.N., K.H., E.M., A.G., J.R.G., K.H., A.A.K., T.D., A.B.-O., S.H.E.K. and S.M.A. contributed to some experiments. L.J., P.B., A.G. and J.R.G. performed the microarray data analysis. T.D., S.H.E.K. and S.M.A. helped with the writing of the manuscript. S.M.A. helped with the design of some experiments. S.F. designed the study, performed some experiments and wrote the manuscript.

Author Information The gene array data have been deposited in NCBI's Gene Expression Omnibus database with the accession number GSE35998. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.F. (fillatreau@drfz.de).

METHODS

Mice, immunization and infection. C57BL/6, $Tlr2^{-/-}$, $Tlr4^{-/-}$, $Ebi3^{-/-}$, $p28^{-/-}$, $p35^{-/-}$, $p40^{-/-}$, $Cd40^{-/-}$, $Il10^{-/-}$, $Il10^{-/-}p35^{-/-}$, $p35^{-/-}Ebi3^{-/-}$p$40^{-/-}$, JHT, IL-10–eGFP[6], and $p35^{-/-}$ IL10–eGFP mice were bred under specific pathogen-free conditions. $B^{Tlr2^{-/-}}$, $B^{Tlr4^{-/-}}$, $B^{p28^{-/-}}$, $B^{p35^{-/-}}$, $B^{Ebi3^{-/-}}$, $B^{p40^{-/-}}$, $B^{Cd40^{-/-}}$ and $B^{Il10^{-/-}}$ mice were obtained by a previously described mixed bone marrow chimaera approach using lethally irradiated C57BL/6 or JHT mice as recipients[3,6]. Briefly, $B^{p35^{-/-}}$ mice were obtained by reconstituting recipient mice with a mixture of bone marrow cells from B-cell-deficient JHT mice (80%) and p35-deficient mice (20%). Control B^{WT} chimaera were obtained using a mixture of bone marrow cells from JHT mice (80%) and C57BL/6 mice (20%) for experiments shown in Fig. 1a, Fig. 2a left panel, and Fig. 2a, right panel. For all other experiments (including Fig. 2a middle panel, and Extended Data Fig. 2a), corresponding B^{WT} control chimaera were obtained as outlined in Extended Data Fig. 10, by reconstituting irradiated mice with a mixture of bone marrow cells from C57BL/6 mice (80%) and from mice deficient in the gene of interest (20%). The various control chimaera are called 'corresponding B^{WT} mice' throughout the study. EAE was induced by immunization with $MOG_{35–55}$ peptide, or recombinant human MOG extracellular domain emulsified in Complete Freund's adjuvant (Sigma-Aldrich) and pertussis toxin (Sigma-Aldrich), and assessed as previously described[3]. Mice with a weight loss >20% were humanely euthanized according to ethical regulations. To measure the MOG-reactive T cell response, 8×10^5 splenocytes were stimulated in flat-bottomed 96-well plates with different concentrations of $MOG_{35–55}$ peptide for 2 days, and culture supernatants were then analysed by ELISA to determine concentrations of IFN-γ and IL-17 (coating/detection with clones R4-6A2 and XMG1.2 for IFN-γ, and TC11-18H10 and TC11-8H4 for IL-17). Mice were infected intravenously with 100 c.f.u. *Salmonella* strain SL1344, and 10^6 or 10^7 c.f.u. attenuated *Salmonella* strain SL7207, and assessed as previously described, including for histological analyses[6]. All experiments were reviewed and approved by appropriate institutional review committees (University of Edinburgh ethical review committee, Comite d'Ethique Midi-Pyrenees, and LAGeSo Berlin), and were conducted in accordance with UK, French and German legislations, in compliance with European community council directive 68/609/EEC guidelines. Mice were of C57BL/6 strain, 6–12-weeks old at the start of experiments, and of male and female genders. EAE experiments and *Salmonella* infections were performed in a blinded manner, and identities of the mice were revealed upon termination of the experiment. No randomization was used. Estimation of size groups was based on our previous experience with these disease models, without a priori determination by power calculation.

B-cell purification and activation. B cells were obtained by magnetic isolation using negative selection with anti-CD11b, anti-CD11c, and anti-CD43 microbeads (Miltenyi Biotec). B cells (>99% pure) were activated as previously described[15] at 5×10^5 cells per well in 96-well flat-bottom plates with LPS (*Escherichia coli* serotype 055:B5; Sigma-Aldrich), CpG-ODN-1826 (TIB MolBiol, Germany), PGN (*Streptomyces* species 79682; Sigma-Aldrich), agonistic anti-CD40 antibody (clone FGK-45, produced in house), mouse CD40L-expressing L47 cells (L47-CD40L$^+$), or control L5 cells (L5-ctrl) as indicated. Culture supernatants were collected at 72 h, and IL-10 concentrations measured by Bio-Plex (Bio-Rad). For the microarray experiments, B cells purified as described above were further depleted of possible contaminants by another round of magnetic negative selection after relabelling with anti-CD11b, anti-CD11c, anti-CD43, anti-CD90, and anti-DX5 microbeads (Miltenyi Biotec). B cells were activated with LPS (1 μg ml^{-1}) or LPS (1 μg ml^{-1}) plus anti-CD40 (10 μg ml^{-1}). Dead cells were eliminated from the 24 h and 72 h activated culture by labelling with propidium iodide (PI), and sorting on FACS Diva (BD Biosciences). Cells were lysed in RLT buffer (Qiagen) and total RNA was extracted using RNeasy Mini Kit (Qiagen).

Gene array hybridization and data analysis. cRNA were hybridized on Affymetrix MG 430 2.0 arrays, using standard Affymetrix protocol after quality control with Agilent 2100 Bioanalyzer and quantification with NanoDrop ND-1000 spectrophotometer, as previously described[29]. The significantly differentially regulated genes were detected using a *t*-test based R-script, with *P* values adjusted using Benjamini Hochberg procedure. In order to be selected in a comparison of two conditions, each Affymetrix ID had to fulfill the following criteria: (1) be present in at least three of the four arrays for at least one of the two conditions compared, (2) to have a mean signal intensity higher than 50 in at least one of the two conditions, and (3) to show an adjusted P value <0.01 (*t*-test) in the comparison of the two conditions. The genes differentially expressed between TLR4-activated and TLR4-activated plus CD40-activated B cells (*t*-test; $P < 0.01$), and differentially modulated during B-cell activation (*t*-test; $P < 0.01$), were then selected and filtered using the gene ontology resource (http://www.geneontology.org) to focus on secreted molecules (Extended Data Fig. 1c). Seven genes fulfilled these criteria, among which five were uniquely increased in TLR4 plus CD40-stimulated B cells (Extended Data Fig. 1c). Hierarchical clustering was performed with the MeV program (version 4.8.1)[30] using Pearson correlation and average linkage.

Analysis of mRNA expression by B and plasma cells. Sorted B and plasma cells were lysed in TRIzol, and RNA was prepared (AMS Biotechnology). After DNase treatment (Ambion), RNA was reverse-transcribed with a Reverse Transcription System (Promega). Quantitative RT–PCR was performed on an MX3005P QPCR System (Stratagene), with LightCycler FastStart DNA Master SYBR Green I (Roche). Transcripts were quantified using β-actin as standard, and the following forward (FP) and reverse (RP) primers (MWG Biotech): β-actin FP: 5′-TGGAATCCTG TGGCATCCATGAAAC-3′, β-actin RP: 5′-TAAAACGCAGCTCAGTAACAG TCC-3′; EBI3 FP: 5′-CGGTGCCCTACATGCTAAAT-3′, EBI3 RP: 5′-GCGGAG TCGGTACTTGAGAG-3′; p35 FP: 5′- CATCGATGAGCTGATGCAGT-3′, p35 RP: 5′-CAGATAGCCCATCACCCTGT-3′; IL-10 FP: 5′-AGCCGGGAAGAC AATAACTG-3′, IL-10 RP: 5′-CATTTCCGATAAGGCTTGG-3′; Blimp1 FP: 5′- GGCATTCTTGGGAACTGTGT-3′; Blimp1 RP: 5′- GACAGAGGCCGAGT TTGAAG-3′; IRF4 FP: 5′- GCAGCTCACTTTGGATGACA-3′; IRF4 RP: 5′-CC AAACGTCACAGGACATTG-3′; Pax5 FP: 5′-AACTTGCCCATCAAGGTGTC-3′; Pax5 RP: 5′-CTGATCTCCCAGGCAAACAT-3′.

Western blot and immunoprecipitation. Activated B cells were treated with GolgiStop (BD Biosciences, Germany) to block protein secretion during the last 4 h of stimulation. B cells (2×10^7) were lysed with 500 μl RIPA buffer (Thermo Fisher Scientific, USA) supplemented with protease inhibitors (Thermo Fisher Scientific, USA). Cells isolated from mice infected with *Salmonella* were directly lysed in RIPA buffer containing protease inhibitors. Protein concentrations of lysates were determined using the BCA Protein Assay Kit (Thermo Fisher Scientific, USA). Proteins were separated on a polyacrylamide gel and transferred to a PVDF membrane (Bio-Rad Laboratories, USA) using semi-dry blotting. EBI3, p35 and actin were detected using rabbit anti-EBI3 (M-75 polyclonal IgG, Santa Cruz Biotechnology, USA), rabbit anti-p35 (EPR5736 polyclonal IgG, Abcam, UK), or rabbit anti-actin (I-19 polyclonal Ab, Santa Cruz Biotechnology, USA) primary antibody, and horseradish peroxidase (HRP)-conjugated secondary anti-rabbit antibody (cat. number: 81-6120, Invitrogen, USA; or cat. number: 111-035-144, Jackson ImmunoResearch, USA) with ECL (GE Healthcare, UK) as HRP substrate. The chemiluminescence signal was measured using the Image-Reader LAS-3000 (Fujifilm, Japan). For immunoprecipitation, supernatant from B cells activated with LPS plus anti-CD40 (clone FGK-45, 10 μg ml^{-1}) were incubated overnight at 4 °C with 2 μg ml^{-1} anti-p35 (C18.2, eBioscience). Immunoprecipitation was performed using μMACS Protein G Microbeads (Miltenyi Biotech), followed by immunoblot to detect EBI3.

B–T cell co-cultures. The protocol for B–T cell co-cultures was adapted from a previous report[25]. Briefly, B cells were magnetically sorted from pooled spleens and lymph nodes of $B^{p35^{-/-}}$ or B^{WT} mice on day 10 post-EAE induction as CD19$^+$ cells (~98% pure), and Teff cells were FACS-sorted from the CD19-depleted fraction as previously described[31]. 50×10^4 B cells and 1×10^4 CD4$^+$CD25$^-$ T cells (Teff) were then co-cultured in the indicated combinations in presence of increasing concentrations of $MOG_{35–55}$. After 48 h cultures received 1 μCi [^{3}H]thymidine, and [^{3}H]thymidine incorporation was measured 16 h later with a Top-Count NXT liquid scintillation counter (Perkin Elmer). Before addition of [^{3}H]thymidine, samples of culture supernatants were collected to quantify concentrations of IL-17, IFN-γ, GM-CSF and IL-6 using Bio-Plex (Bio-Rad).

Plasma cell purification. Plasma cells and B cells were obtained from C57BL/6, $p35^{-/-}$, $p35^{-/-}Ebi3^{-/-}p40^{-/-}$, IL-10–eGFP, and p35$^{-/-}$IL-10–eGFP mice on day 3 after infection with 10^7 colony forming units (c.f.u.) *Salmonella* (SL7207) by magnetic isolation using anti-CD138-PE (clone 281-2, BD Pharmingen) and anti-PE microbeads (Miltenyi Biotec). The negative fraction was then subjected to FACS sorting to obtain high purity CD19$^+$CD138$^-$ B cells. The positive fraction was then stained for CD22 (clone OX-97, BioLegend), and subjected to FACS sorting to obtain high purity plasma cells (CD138^{hi}), and plasma cell subsets (CD138^{int}CD22$^+$, CD138^{hi}CD22$^+$ and CD138^{hi}CD22$^-$ cells). For western blot, CD138^{hi} and CD138^{int} cells were isolated as CD138$^+$ plasma cells from infected mice by magnetic isolation using anti-CD138-PE (clone 281-2, BD Pharmingen) and anti-PE microbeads (Miltenyi Biotec).

Single-cell PCR analysis. Single cells were sorted on a FACS Aria II (BD Biosciences) into a 96-well PCR plate, immediately frozen in liquid nitrogen and stored at −80 °C until further use. For detection of respective transcripts a two-step PCR approach was used. Reverse transcription and the first PCR step were carried out in a one-step reaction using the Qiagen OneStep RT–PCR kit according to the manufacturer's instructions. As recommended in these instructions, specific nested primers (MWG Biotech) were used as follows: EBI3nested FP: 5′-CCTTCATT GCCACTTACAGG-3′, EBI3nested RP: 5′-TAATCTGTGAGGTCCTGAGC-3′; p35nested FP: 5′-CATTCTAGACAAGGGCATGC-3′, p35nested RP: 5′-GTGAT GGGAGAACAGATTCC-3′; IL-10nested FP: 5′-TCTTACTGACTGGCATGAG G-3′; IL-10nested RP: 5′-CTTCTACCAGGTAAAACTGG-3′; Blimp1nested FP:

5′- CGTGAAGTTTCAAGGACTGG-3′; Blimp1nested RP: 5′- GTGGTGGAAC TCCTCTCTGG-3′. For validation of sorting, β-actin primers were added to the reaction mixture. After this first reaction, an aliquot of the PCR product was loaded on an agarose gel, and only β-actin positive samples were considered to further analysis. A 100-fold dilution of the PCR product was subsequently used as template for the second PCR reaction using the primers described in the section: analysis of mRNA expression by B and plasma cells. Amplification of the respective transcript was verified on an agarose gel.

Histology. For immunostaining, 1- to 2-μm sections of formalin-fixed, paraffin-embedded tissue were cut, deparaffinized and subjected to a heat-induced epitope retrieval step. Slides were rinsed in cool running water and washed in Tris-buffered saline (pH 7.4) before incubation with primary antibodies against CD3 (cat number N1580, Dako, Glostrup, Denmark, dilution 1:10), and F4/80 (clone BM8, eBioscience, dilution 1:50) for 30 min, followed by biotinylated donkey anti-rat or donkey anti-rabbit (cat. numbers 712-065-153 and 711-065-152, Dianova, Hamburg, Germany) secondary antibodies. The streptavidin-AP kit (cat. number K5005, Dako) was used for detection. As negative control, primary antibody was omitted.

Detection of serum antibodies by ELISA. To quantify Salmonella-specific or NP-OVA-specific antibodies, diluted sera were incubated on 96-well plates coated with 5×10^6 heat-killed *S. typhimurium* per well or $50\ \mu g\ ml^{-1}$ NP(15)-BSA (Biosearch Technologies), respectively. Detection was done with alkaline phosphatase-conjugated anti-IgM (cat number 1020-04), anti-IgG (cat number 1030-04), anti-IgG1 (cat number 1070-04), anti-IgG2b (cat number 1090-04), anti-IgG2c (cat number 1079-04), and anti-IgG3 (cat number 1100-04) antibodies (all from Southern Biotechnology Associates).

Detection of mouse IL-27 p28 subunit by ELISA. The concentration of mouse IL-27 p28 subunit was quantified using the mouse IL-27 p28 Quantikine ELISA Kit (cat number M2728, R&D Systems), according to the manufacturer's instructions.

Isolation of CNS-infiltrating cells. CNS single-cell suspension was obtained by collagenase digestion of brain and spinal cord. Infiltrating leukocytes were separated from debris and tissue cells using Ficoll centrifugation. Cell numbers were determined by FACS. Cells were characterized by flow cytometry after staining for CD4 (clone RM4-5), CD11b (clone M1/70), Ly6G (clone 1A8) and Foxp3 (clone FJK16 s).

Treg suppression assay. $CD4^+CD25^+$ T cells (Treg) were isolated from $B^{p35-/-}$ and B^{WT} mice on day 10 post-EAE induction, and compared for their capacity to suppress the proliferation of $CD4^+CD25^-$ T cells (Teff) isolated from naive C57BL/6 mice *in vitro*, as previously described[31].

ELISPOT Assay. Sorted B cells ($CD19^+CD138^-$), and plasma cell subsets ($CD138^{int}CD22^+$, $CD138^{hi}CD22^+$, and $CD138^{hi}CD22^-$ cells) or total plasma cells ($CD138^+$) were seeded at a starting number of 10^4 cells per well, with seven successive threefold serial dilutions, in 96-well flat-bottom EIA/RIA high-binding plate (Millipore) pre-coated with anti-mouse Ig(H+L) chain ($5\ \mu g\ ml^{-1}$; Southern Biotechnology Associates, cat number 1010-01). After 3 h incubation, plates were washed, and incubated with alkaline phosphatase-conjugated anti-IgM or anti-IgG antibodies overnight at 4 °C (Southern Biotechnology Associates, cat numbers 1020-04 and 1030-04). ELISPOT were then developed using BCIP/NBT substrate (Gene Tex Inc.).

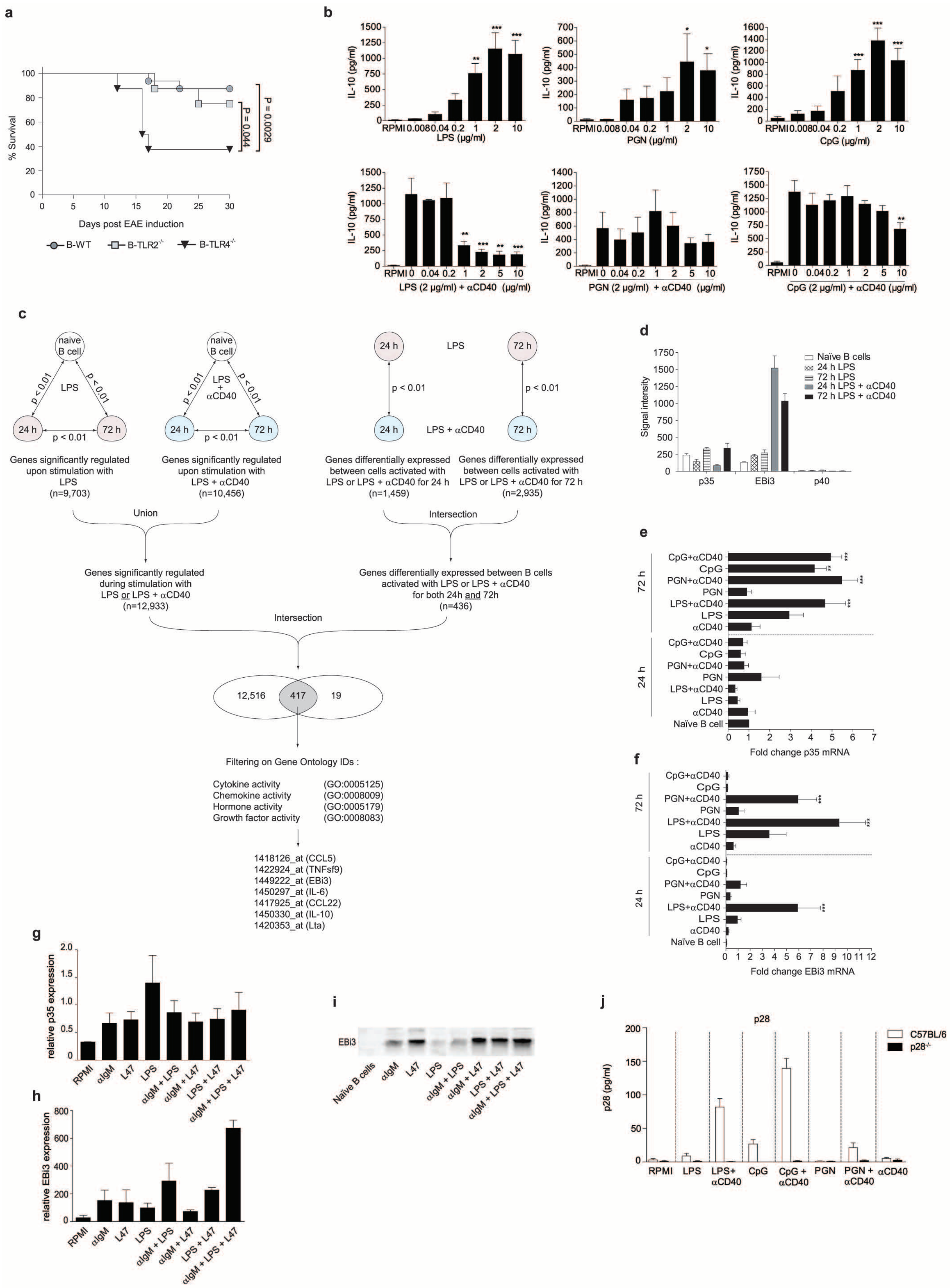
Stimulation and characterization of plasma cells and B cells. Purified plasma cells and B cells were seeded at 5×10^5 cells per well on 96-well flat-bottom plates and stimulated with PMA plus ionomycin, LPS ($1\ \mu g\ ml^{-1}$; *Escherichia coli* serotype 055:B5; Sigma-Aldrich), agonistic anti-CD40 antibody ($10\ \mu g\ ml^{-1}$; clone FGK-45, produced in house), anti-IgM ($5\ \mu g\ ml^{-1}$, cat number 115-006-075, Jackson ImmunoResearch Laboratories, USA), IL-4 ($20\ ng\ ml^{-1}$; R & D Systems GmbH), IL-5 ($20\ ng\ ml^{-1}$; eBioscience), IL-6 ($20\ ng\ ml^{-1}$; R & D Systems GmbH), IL-21 ($20\ ng\ ml^{-1}$; R & D Systems GmbH), and indicated combinations. After 24 h, supernatants were harvested for the detection of IL-10 and IL-6 concentration by Bio-Plex (Bio-Rad, USA); cells were subjected to flow cytometry for the quantification of GFP-positive cells. Characterizations were done by flow cytometry after staining for CD19 (clone 1D3, BD Pharmingen), CD138 (clone 281-2, BD Pharmingen), IgM (cat number 1020-02, SouthernBiotech), IgD (clone 217-170, BD Pharmingen), CD80 (clone 16-10A1, BD Pharmingen), CD86 (clone GL1, BD Pharmingen), MHCII (clone M5/114, in house), CD40 (clone 3/23, BD Pharmingen), CD69 (clone H1.2F3, eBioscience), CD44 (clone 1M7, in house), CD43 (clone S7, BD Pharmingen), CD28 (clone E18, Biolegend), TACI (clone 8F10, BD Pharmingen), CXCR4 (clone 2B11, BD Pharmingen), CD1d (clone 1.B1, in house) and Tim1 (clone RMT1-4, Biolegend).

Validation of chimaeras. $B^{Ebi3-/-}$ and corresponding B^{WT} chimaeras were used as an example to demonstrate the validity of the chimaera system, that is, that the deficiency in the gene of interest is restricted to B cells, and complete in that compartment in $B^{X-/-}$ mice. To this end, B cells, $CD11b^+$ cells and $CD4^+$ T cells were purified from $B^{Ebi3-/-}$ and corresponding B^{WT} chimaeras. Briefly, splenocytes from $B^{Ebi3-/-}$ and B^{WT} mice were subjected to CD19 positive selection using magnetic cell sorting. The $CD19^+$ fraction was subsequently stained for B220 (clone RA3-6B2), CD11b (clone M1/70), CD11c (clone N418), CD4 (clone GK1.5), and CD8 (clone 53.6.72), and FACS-sorted for $B220^+CD11b^-CD11c^-CD4^-CD8^-$ cells, yielding >99% pure B cells. The $CD19^-$ fraction obtained from the magnetic sort was stained, and used to FACS-sort $CD4^+B220^-CD11b^-$ cells, and $CD11b^+B220^-CD4^-$ cells, resulting in >97–98% pure $CD4^+$ T cells and $CD11b^+$ cells, respectively. All sorted populations as well as the remaining non-T/B/$CD11b^+$ cells (denoted as non-T, B, M cells) were lysed and genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen), according to the manufacturer's instructions. The amounts of genomic DNA present in these samples were equilibrated by real time PCR using the *Mog* gene as standard with the primer pair: forward 5′-AGGAAGGGACATGCAGCCGGA-3′; reverse 5′-CTGC ATAGCTGCATGACAACTG-3′. Wild-type *Ebi3* gene was then amplified by conventional PCR with the following primers: forward 5′-AACCTCAGGCCA GGCAGT-3′; reverse 5′-TTCCGTAGGCCATGTAGGAC-3′, in order to test for the presence of the wild-type *Ebi3* allele in the different cell fractions.

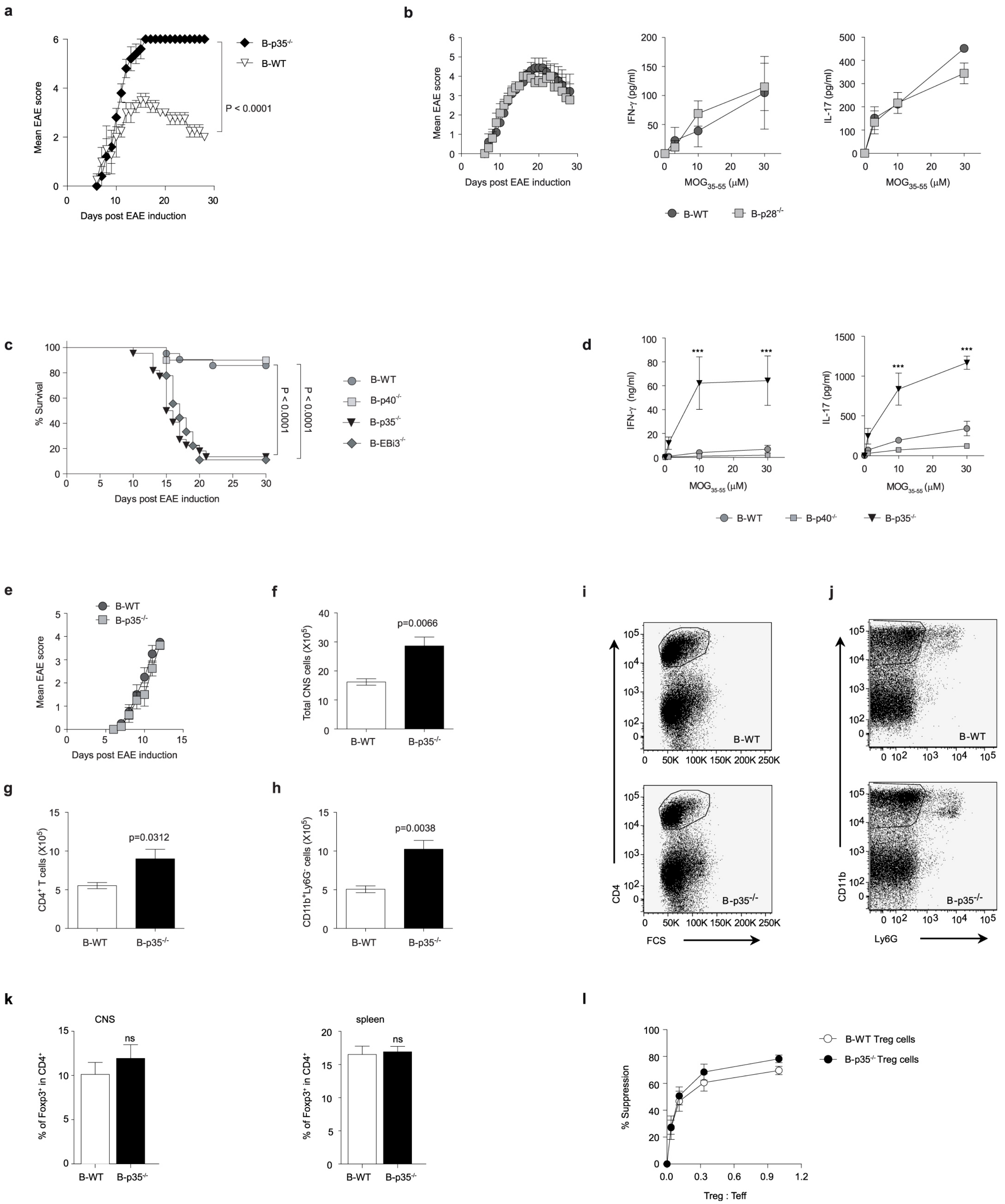
Additional reagents. Antibodies used in this study also included anti-Ly6C (clone AL21, cat number 557359, BD Pharmingen), anti-IFN-γ (clone XMG1.2, BD Pharmingen), anti-CD154 (cat number 130-092-105, Miltenyi Biotech), anti-CD62L (MEL14, in house).

Statistics. Statistical analysis was performed using GraphPad Prism (version 5.02 for Windows, GraphPad Software, USA). EAE data distribution did not differ from normal distribution, as evaluated using Kolmogorov–Smirnov test. Equality of variances between groups was assessed before analyses by ANOVA or *t*-test. Groups were compared using ANOVA, two-tailed *t*-test, or Wilcoxon test, as indicated in figure legends. The *t*-tests were modified using Welch's correction in case of unequal variance. One-way and two-way ANOVA were followed by Bonferroni post-test. No samples were excluded from analysis. Statistical analysis of the gene array data are described in the section gene array hybridization and data analysis.

29. Biesen, R. *et al.* Sialic acid-binding Ig-like lectin 1 expression in inflammatory and resident monocytes is a potential biomarker for monitoring disease activity and success of therapy in systemic lupus erythematosus. *Arthritis Rheum.* **58,** 1136–1145 (2008).
30. Saeed, A. I. *et al.* TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34,** 374–378 (2003).
31. Hoehlig, K. *et al.* Activation of $CD4^+$ $Foxp3^+$ regulatory T cells proceeds normally in the absence of B cells during EAE. *Eur. J. Immunol.* **42,** 1164–1173 (2012).

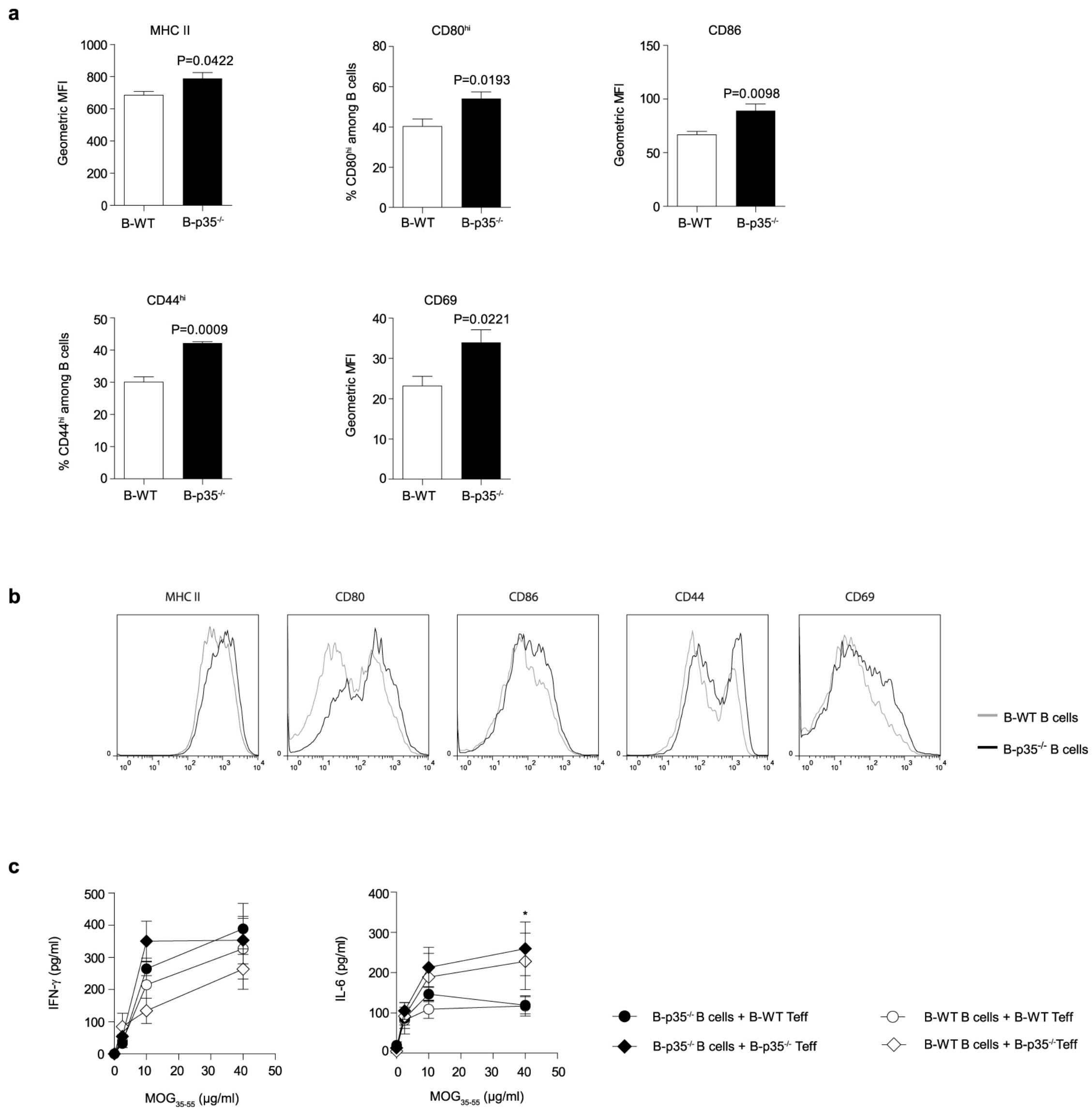

a
b
c
d
e
f
g
h
i
j
% Survival
Days post EAE induction
B-WT
B-TLR2-/-
B-TLR4-/-
P = 0.0029
P = 0.044
IL-10 (pg/ml)
LPS (µg/ml)
PGN (µg/ml)
CpG (µg/ml)
LPS (2 µg/ml) + αCD40 (µg/ml)
PGN (2 µg/ml) + αCD40 (µg/ml)
CpG (2 µg/ml) + αCD40 (µg/ml)
naive B cell
LPS
LPS + αCD40
p < 0.01
Genes significantly regulated upon stimulation with LPS (n=9,703)
Genes significantly regulated upon stimulation with LPS + αCD40 (n=10,456)
Genes differentially expressed between cells activated with LPS or LPS + αCD40 for 24 h (n=1,459)
Genes differentially expressed between cells activated with LPS or LPS + αCD40 for 72 h (n=2,935)
Union
Intersection
Genes significantly regulated during stimulation with LPS or LPS + αCD40 (n=12,933)
Genes differentially expressed between B cells activated with LPS or LPS + αCD40 for both 24h and 72h (n=436)
12,516
417
19
Filtering on Gene Ontology IDs :
Cytokine activity (GO:0005125)
Chemokine activity (GO:0008009)
Hormone activity (GO:0005179)
Growth factor activity (GO:0008083)
1418126_at (CCL5)
1422924_at (TNFsf9)
1449222_at (EBi3)
1450297_at (IL-6)
1417925_at (CCL22)
1450330_at (IL-10)
1420353_at (Lta)
Signal intensity
Naïve B cells
24 h LPS
72 h LPS
24 h LPS + αCD40
72 h LPS + αCD40
p35
EBi3
p40
Fold change p35 mRNA
Fold change EBi3 mRNA
relative p35 expression
relative EBi3 expression
EBi3
p28
p28 (pg/ml)
C57BL/6
p28-/-

Extended Data Figure 1 | Gene array analysis and IL-35 secretion by B cells activated *in vitro*. **a**, EAE was induced in $B^{Tlr2^{-/-}}$ (squares), $B^{Tlr4^{-/-}}$ (triangles) and B^{WT} (circles) mice (see Fig. 1). Animals showing a body weight loss >20% and/or sustained front leg impairment were killed in accordance with ethical regulations. Survival curves were compared using Wilcoxon test. **b**, In the top panels, splenic B cells from C57BL/6 mice were stimulated with lipopolysaccharides (LPS), peptidoglycan (PGN) or CpG DNA oligonucleotides for 72 h, as indicated, and IL-10 concentrations in supernatants were determined by Bio-Plex. Results from stimulated cells were compared to unstimulated B cells (RPMI) using one-way ANOVA followed by Bonferroni post-test (***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; mean ± s.e.m.); in the bottom panels, splenic B cells were simultaneously activated with anti-CD40 (clone FGK-45; in indicated amounts) and LPS (2 $\mu g\,ml^{-1}$), PGN (2 $\mu g\,ml^{-1}$) or CpG (2 $\mu g\,ml^{-1}$). Supernatants were collected after 72 h and analysed for IL-10 content by Bio-Plex. Data shown are pooled from three independent experiments (mean ± s.e.m.). Results from B cells stimulated with LPS, CpG or PGN alone were compared to B cells co-stimulated with anti-CD40 using one-way ANOVA followed by Bonferroni post-test (***$P < 0.001$; **$P < 0.01$; *$P < 0.05$). **c**, Affymetrix MG 430 2.0 whole genome arrays were performed in quadruplicates for naive B cells, B cells activated by TLR4 for 24 h and 72 h, and B cells activated by TLR4 plus CD40 for 24 h and 72 h (20 arrays in total). To obtain genes significantly regulated upon stimulation with LPS, the expression profiles of naive B cells, B cells activated with LPS for 24 h, and B cells activated with LPS for 72 h were compared to each other, generating three lists of differentially expressed Affymetrix IDs (referred to as genes in this figure), whose union provided a set of $n = 9{,}703$ Affymetrix IDs differentially expressed during LPS activation. A similar analysis yielded a set of $n = 10{,}456$ Affymetrix IDs differentially expressed during LPS plus anti-CD40 activation. The union of these two sets was computed to obtain the list of Affymetrix IDs modulated during B-cell stimulation with LPS or LPS plus anti-CD40 ($n = 12{,}933$). To focus on genes differentially expressed between B cells activated with LPS or LPS plus anti-CD40, the expression profiles of B cells activated with LPS or LPS plus anti-CD40 for 24 h were compared to each other. A similar analysis was done for B cells activated for 72 h. The intersection of these two sets provided the list of Affymetrix IDs differentially expressed between LPS-activated and LPS-activated plus anti-CD40-activated B cells at both 24 h and 72 h (this was done to identify the Affymetrix IDs with long-term differential expression) ($n = 436$). The intersection of the set $n = 12{,}933$ with the set $n = 436$ gave a list of 417 Affymetrix IDs, which was further filtered on Gene Ontology to extract the Affymetrix IDs corresponding to secreted factors, yielding a final list of 7 genes. **d**, Signal intensities of *p35*, *Ebi3* and *p40* mRNA expression were calculated using the values of the Affymetrix arrays. Data show mean ± s.e.m. **e**, **f**, Splenic murine B cells were stimulated with LPS (1 $\mu g\,ml^{-1}$), PGN (10 $\mu g\,ml^{-1}$), CpG ODN 1826 (1 $\mu g\,ml^{-1}$), and anti-CD40 (clone FGK-45; 10 $\mu g\,ml^{-1}$), alone or in combination, as indicated, for 24 h or 72 h. *p35* (**e**) and *Ebi3* (**f**) mRNA expression was then quantified by real-time PCR. Data compile three independent experiments (mean ± s.e.m.). For statistical analysis, one-way ANOVA test followed by Bonferroni post-test was performed comparing each activated B-cell sample to naive B cells (***$P < 0.001$; **$P < 0.01$; mean ± s.e.m.). **g**–**i**, Splenic murine B cells were stimulated for 72 h with anti-IgM (F(ab′)2 goat anti-mouse IgM; Jackson ImmunoResearch; 5 $\mu g\,ml^{-1}$), LPS (1 $\mu g\,ml^{-1}$) and L47 cells (5×10^4 cells, irradiated), alone or in combinations, as indicated. **g**, **h**, Levels of *p35* and *Ebi3* mRNA expression were quantified by real-time PCR. Data show compilation of three independent experiments (mean ± s.e.m.). **i**, Splenic murine B cells were activated as indicated for 72 h, and treated with GolgiStop for the last 4 h of culture. B-cell lysates were separated on SDS–PAGE gel and blotted with anti-EBI3 antibody. Data show representative results from three independent experiments. **j**, Splenic B cell from C57BL/6 and *p28* mice were stimulated with LPS (1 $\mu g\,ml^{-1}$), CpG (1 $\mu g\,ml^{-1}$), PGN (10 $\mu g\,ml^{-1}$), anti-CD40 (10 $\mu g\,ml^{-1}$), alone or in combination, as indicated, for 72 h. p28 concentrations in culture supernatants were determined by ELISA. Data shown (mean ± s.em) are pooled from four independent experiments.

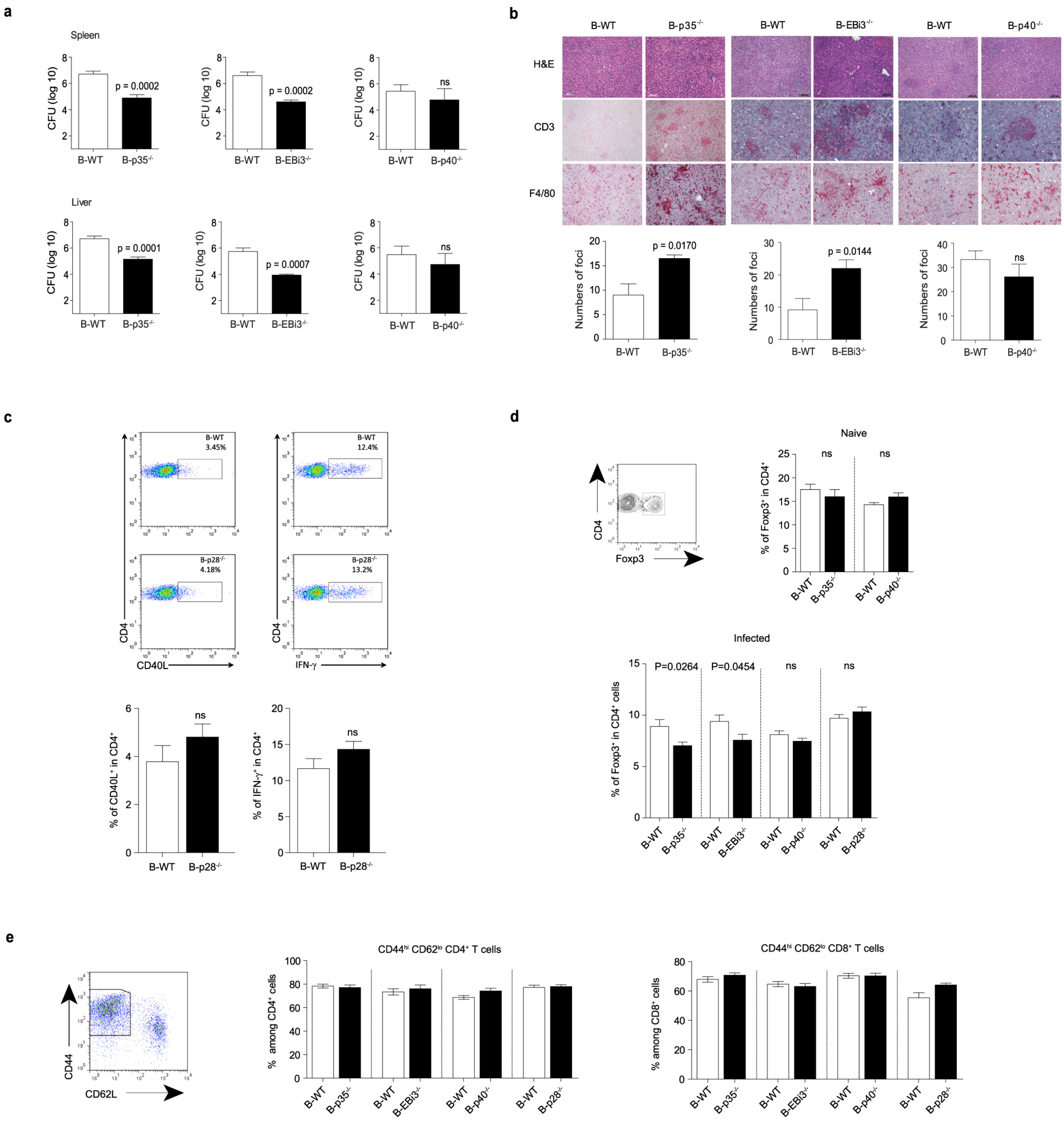


a
Mean EAE score
Days post EAE induction
B-p35-/-
B-WT
P < 0.0001
b
Mean EAE score
Days post EAE induction
IFN-γ (pg/ml)
IL-17 (pg/ml)
MOG35-55 (μM)
B-WT
B-p28-/-
c
% Survival
Days post EAE induction
P < 0.0001
P < 0.0001
B-WT
B-p40-/-
B-p35-/-
B-EBi3-/-
d
IFN-γ (ng/ml)
IL-17 (pg/ml)
MOG35-55 (μM)
***
B-WT
B-p40-/-
B-p35-/-
e
B-WT
B-p35-/-
Mean EAE score
Days post EAE induction
f
Total CNS cells (X10^5)
p=0.0066
B-WT
B-p35-/-
g
CD4+ T cells (X10^5)
p=0.0312
B-WT
B-p35-/-
h
CD11b+Ly6G- cells (X10^5)
p=0.0038
B-WT
B-p35-/-
i
CD4
FCS
B-WT
B-p35-/-
j
CD11b
Ly6G
B-WT
B-p35-/-
k
CNS
spleen
% of Foxp3+ in CD4+
ns
B-WT
B-p35-/-
l
% Suppression
Treg : Teff
B-WT Treg cells
B-p35-/- Treg cells

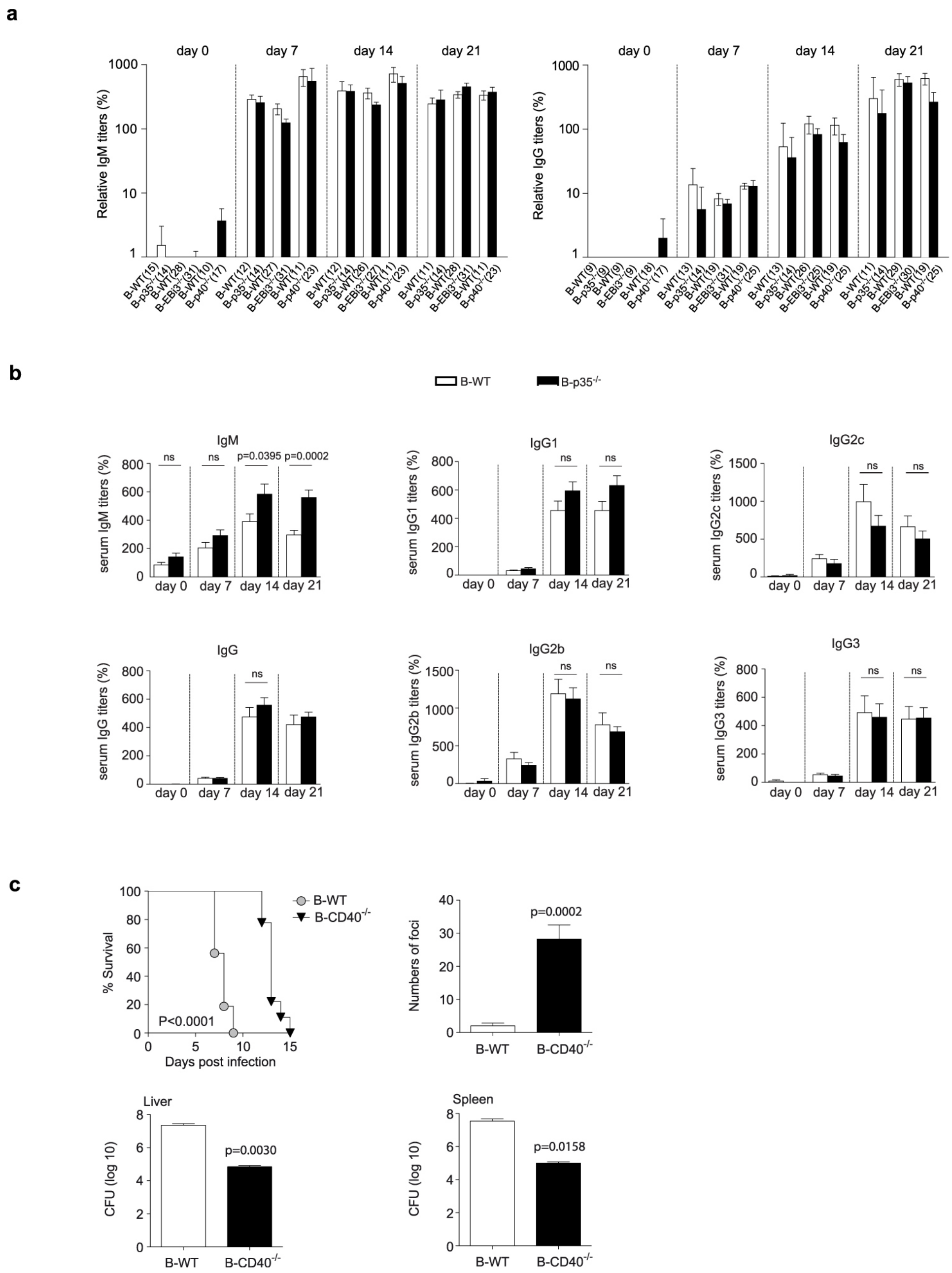
Extended Data Figure 2 | Characterization of immune reponses during EAE in mice lacking IL-35 expression in B cells. **a**, EAE was induced in $B^{p35^{-/-}}$ (black diamonds, $n = 5$) and corresponding B^{WT} mice (white triangles, $n = 4$). Data show clinical EAE scores (mean ± s.e.m.) from two independent experiments. Cumulative disease scores were compared using two-tailed unpaired *t*-test. **b**, EAE was induced in $B^{p28^{-/-}}$ (grey squares; $n = 10$) and corresponding B^{WT} mice (dark grey circles, $n = 10$). Splenocytes were collected from mice on day 28 after EAE induction, and re-stimulated individually for 48 h with MOG_{35-55}. Supernatants were analysed by ELISA to determine concentrations of IFN-γ (middle) and IL-17 (right). Data (mean ± s.e.m.) are pooled from two independent experiments. **c**, EAE was induced in $B^{p35^{-/-}}$ (triangles), $B^{Ebi3^{-/-}}$ (diamonds), $B^{p40^{-/-}}$ (squares), and B^{WT} (circles) mice (see Fig. 2). Animals showing a body weight loss > 20% and/or sustained front leg impairment were killed in accordance with ethical regulations. Survival curves were compared using Wilcoxon test. **d**, Splenocytes were collected from $B^{p35^{-/-}}$, $B^{p40^{-/-}}$, and B^{WT} mice on day 28 after EAE induction (see Fig. 2a), and pooled before re-stimulation for 48 h with MOG_{35-55}. Culture supernatants were analysed by ELISA to determine concentrations of IFN-γ (left) and IL-17 (right). Data show representative results from two independent experiments. Statistical analysis was performed using the two-way-ANOVA test followed by a Bonferroni post-test (*$P < 0.05$, **$P < 0.01$; ***$P < 0.001$; mean ± s.e.m.). Results of analysis are shown for B^{WT} versus $B^{p35^{-/-}}$ comparison. **e**–**j**, EAE was induced in $B^{p35^{-/-}}$ (grey squares; $n = 8$) and corresponding B^{WT} mice (dark grey circles; $n = 8$) and mice were killed at day 12 post-EAE induction to analyse accumulation of immune cells in CNS. Data show clinical EAE scores (mean ± s.e.m.) (**e**). **f**–**h**, Mononuclear cells were isolated from CNS (brain plus spinal cord) of individual mice shown in (**e**), and numbers of total cells (**f**), $CD4^+$ T cells (**g**) and $CD11b^+Ly6G^-$ macrophages (**h**) were quantified by flow cytometry. Data shown (mean ± s.e.m.) are pooled from two independent experiments. Results were compared using two-tailed unpaired *t*-test with Welch's correction in case of inequal variances. **i**, **j**, Representative FACS plots of $CD4^+$ T cells and $CD11b^+Ly6G^-$ macrophages, respectively. **k**, Frequencies of $Foxp3^+$ Treg cells among $CD4^+CD8^-$ T cells in CNS at day 12 post-EAE induction (left panel), and in spleen at day 10 post-EAE induction (right panel) as determined by flow cytometry. Data shown in the left panel (mean ± s.e.m.) are pooled from two independent experiments with 10 mice per group in total. Data shown in the right panel (mean ± s.e.m.) are representative of three independent experiments with 9 mice per group in total. Results were compared using two-tailed unpaired *t*-test. *P* values >0.05 are indicated by ns. **l**, Treg suppression assay. $CD4^+CD25^+$ T cells (Treg) were FACS-sorted from pooled spleen and lymph nodes of $B^{p35^{-/-}}$ and corresponding B^{WT} mice on day 10 post-EAE induction, and then co-cultured at the indicated ratios with 2×10^4 $CD4^+CD25^-$ T cells (Teff) isolated from naive C57BL/6 mice in presence of 1×10^5 irradiated splenocytes and 0.1 μg ml^{-1} anti-CD3 (clone 145-2C11). Proliferation was measured after 64 h of culture by incorporation of $[^3H]$thymidine, and plotted as suppression percentage defining as 100% the proliferation values obtained for control cultures without Treg cells. Data (mean ± s.e.m.) are pooled from two independent experiments.



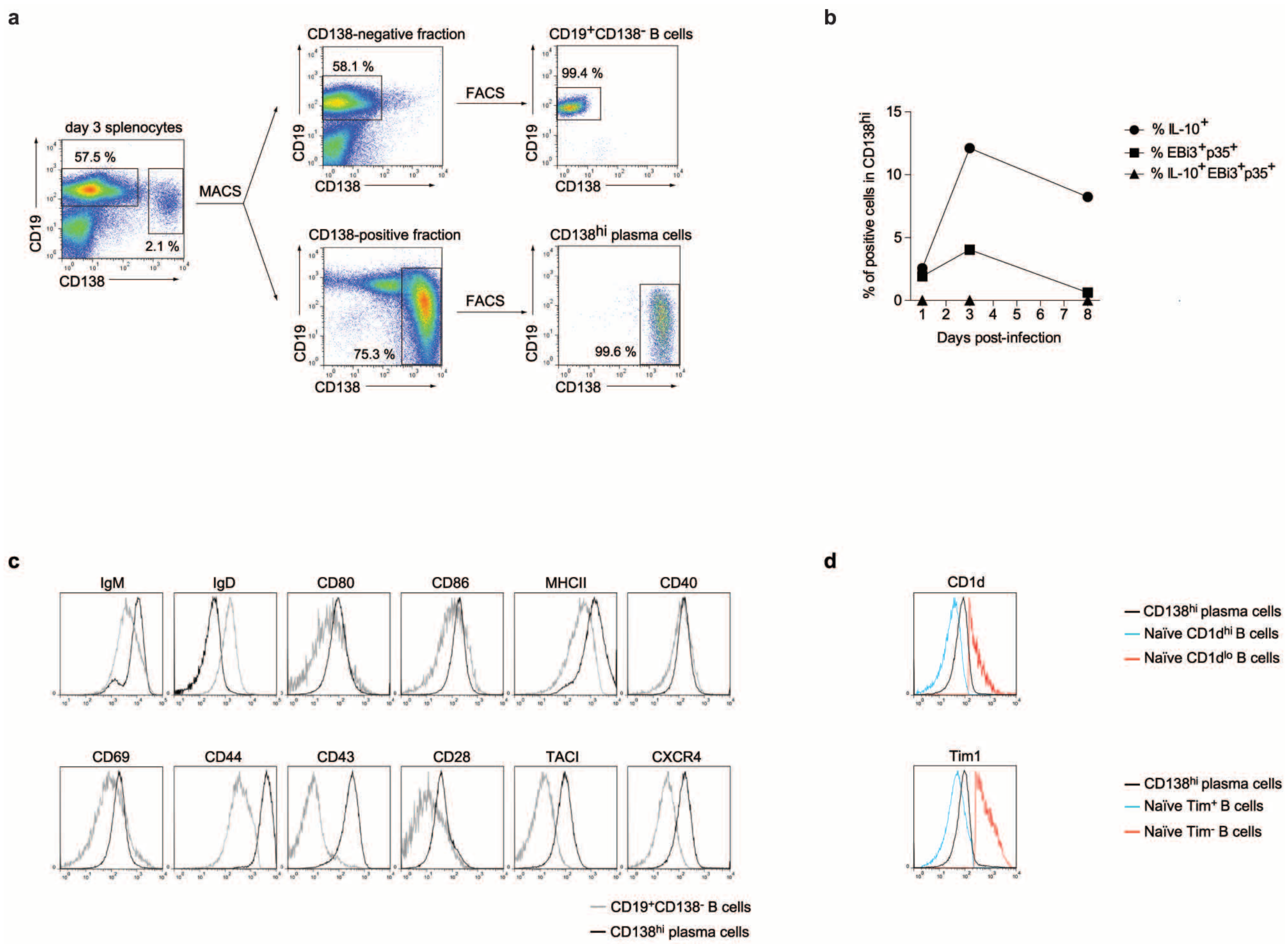
Extended Data Figure 3 | Lack of IL-35 production by B cells results in increased activation of B cells during EAE. **a**, EAE was induced in $B^{p35^{-/-}}$ ($n = 6$) and corresponding B^{WT} mice ($n = 6$). Mice were killed on day 10 after immunization and splenic B cells were analysed by flow cytometry for surface expression of antigen-presenting molecules (MHC II), co-stimulatory molecules (CD80, CD86), and activation markers (CD44, CD69). MFI, mean fluorescence intensity. Data (mean ± s.e.m.) are pooled from two independent experiments. Results were compared using two-tailed unpaired *t*-test with Welch's correction in case of inequal variances. **b**, Representative histogram plots showing expression of these molecules on wild-type and p35-deficient B cells. Dead cells were excluded using propidium iodide. **c**, EAE was induced in $B^{p35^{-/-}}$ and corresponding B^{WT} mice. B cells and $CD4^+CD25^-$ T cells (Teff) were isolated from pooled lymph nodes and spleens on day 10 after immunization. To test the APC function of B cells (see Fig. 2), 5×10^5 B cells from $B^{p35^{-/-}}$ or B^{WT} mice were then cultured with 1×10^4 Teff cells from $B^{p35^{-/-}}$ or B^{WT} mice in presence of MOG_{35-55}, as indicated. Culture supernatants were collected after 48 h, and analysed by Bio-Plex to determine concentrations of IFN-γ (left) and IL-6 (right). Data (mean ± s.e.m.) are pooled from two independent experiments. Results were compared using two-way ANOVA followed by Bonferroni post-test (*$P < 0.05$). Results of analysis are shown for comparison, B^{WT} B cells plus B^{WT} T cells versus $B^{p35^{-/-}}$ B cells plus $B^{p35^{-/-}}$ T cells for IL-6.


a
Spleen
Liver
CFU (log 10)
B-WT
B-p35-/-
B-EBi3-/-
B-p40-/-
p = 0.0002
p = 0.0001
p = 0.0007
ns
b
H&E
CD3
F4/80
Numbers of foci
p = 0.0170
p = 0.0144
c
CD4
CD40L
IFN-γ
B-p28-/-
3.45%
12.4%
4.18%
13.2%
% of CD40L+ in CD4+
% of IFN-γ+ in CD4+
d
Foxp3
Naive
Infected
% of Foxp3+ in CD4+
% of Foxp3+ in CD4+ cells
P=0.0264
P=0.0454
e
CD44
CD62L
CD44hi CD62lo CD4+ T cells
CD44hi CD62lo CD8+ T cells
% among CD4+ cells
% among CD8+ cells

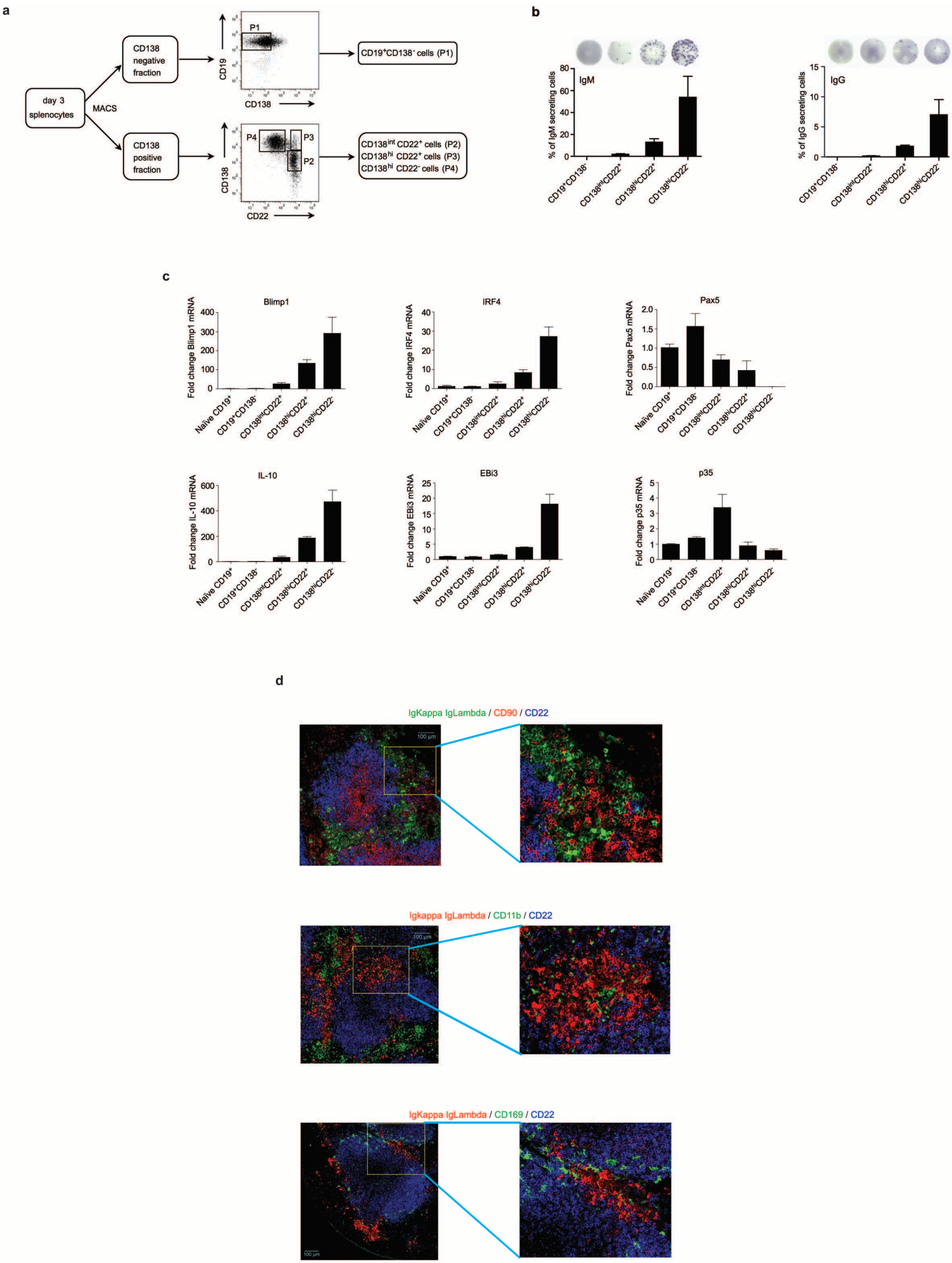
Extended Data Figure 4 | Bacterial burden and T-cell activation in mice lacking IL-35 production by B cells during *Salmonella* infection. **a**, $B^{p35-/-}$, $B^{Ebi3-/-}$, $B^{p40-/-}$ and corresponding B^{WT} mice were infected intravenously with virulent *Salmonella* strain SL1344 (100 c.f.u.). Bacterial loads in spleens (top) and livers (bottom) were determined on day 6 post-infection (p.i). Data (mean ± s.e.m.) are pooled from two independent experiments. Numbers of mice analysed: $B^{p35-/-}$ ($n = 6$) and corresponding B^{WT} mice ($n = 8$), $B^{Ebi3-/-}$ ($n = 5$) and corresponding B^{WT} mice ($n = 5$), $B^{p40-/-}$ ($n = 6$) and corresponding control B^{WT} mice ($n = 5$). Data were analysed with two-tailed unpaired *t*-test. P values > 0.05 are indicated by ns. **b**, Top shows liver histochemistry of total infiltrating cells (haematoxylin and eosin; H&E), T cells (CD3), and macrophages (F4/80) at day 6 p.i.; bottom, numbers of inflammatory foci per liver section on day 6 p.i. Data (mean ± s.e.m.) are pooled from two independent experiments. Numbers of mice analysed: $B^{p35-/-}$ ($n = 6$) and corresponding B^{WT} mice ($n = 8$), $B^{Ebi3-/-}$ ($n = 7$) and corresponding B^{WT} mice ($n = 5$), $B^{p40-/-}$ ($n = 7$) and corresponding B^{WT} mice ($n = 8$). Data were compared using two-tailed unpaired *t*-test with Welch's correction in case of inequal variances. P values > 0.05 are indicated by ns. **c**, $B^{p28-/-}$ ($n = 6$) and corresponding B^{WT} ($n = 6$) mice were infected intravenously with attenuated *Salmonella* (strain SL7207). On day 21 p.i. bone marrow (BM) cells were isolated from individual mice, and re-stimulated for 6 h with heat-killed *Salmonella* before staining for surface CD4, and intracellular CD40L or IFN-γ. (top). Representative FACS plots showing CD40L and IFN-γ expression by $CD4^+$ T cells, after gating on $CD4^+$ cells. Bottom, frequencies of $CD40L^+$ (left) and IFN-γ^+ (right) cells among $CD4^+$ T cells in BM (mean ± s.e.m.). Data are pooled from two independent experiments. Data were compared using two-tailed unpaired *t*-test. P values > 0.05 are indicated by ns. **d**, Representative FACS plot shows Foxp3 expression by splenic $CD4^+$ T cells from a B^{WT} mouse on day 21 p.i. with attenuated *Salmonella* (strain SL7207); (top right) frequency of $Foxp3^+$ Treg cells among $CD4^+$ T cells in spleens of naive $B^{p35-/-}$ ($n = 5$) and corresponding B^{WT} mice ($n = 5$), $B^{p40-/-}$ ($n = 7$) and corresponding B^{WT} mice ($n = 7$); (bottom) frequency of $Foxp3^+$ Treg cells among $CD4^+$ T cells on day 21 p.i. with attenuated *Salmonella* (strain SL7207) in spleens of $B^{p35-/-}$ ($n = 8$) and corresponding B^{WT} mice ($n = 8$), $B^{Ebi3-/-}$ ($n = 10$) and corresponding B^{WT} mice ($n = 10$), $B^{p40-/-}$ ($n = 15$) and corresponding B^{WT} mice ($n = 9$), $B^{p28-/-}$ ($n = 6$) and corresponding B^{WT} mice ($n = 6$). Data (mean ± s.e.m.) are pooled from two independent experiments. Data were analysed with two-tailed unpaired *t*-test. P values >0.05 are indicated by ns. **e**, Mice were infected as in (**d**) and frequency of activated $CD44^{hi}CD62L^{lo}$ $CD4^+$ and $CD8^+$ T cells were determined by flow cytometry. Representative FACS plot shows expression of CD44 and CD62L by live (PI negative) $CD4^+$ T cells. Graphs show frequencies of $CD44^{hi}CD62L^{lo}$ cells among $CD4^+$ T cells (left) and $CD8^+$ T cells (right). Data (mean ± s.e.m.) are pooled from two independent experiments. Numbers of mice analysed: $B^{p35-/-}$ ($n = 14$) and corresponding B^{WT} mice ($n = 12$), $B^{Ebi3-/-}$ ($n = 10$) and corresponding B^{WT} mice ($n = 10$), $B^{p40-/-}$ ($n = 15$) and corresponding B^{WT} mice ($n = 9$), $B^{p28-/-}$ ($n = 6$) and corresponding B^{WT} mice ($n = 6$).



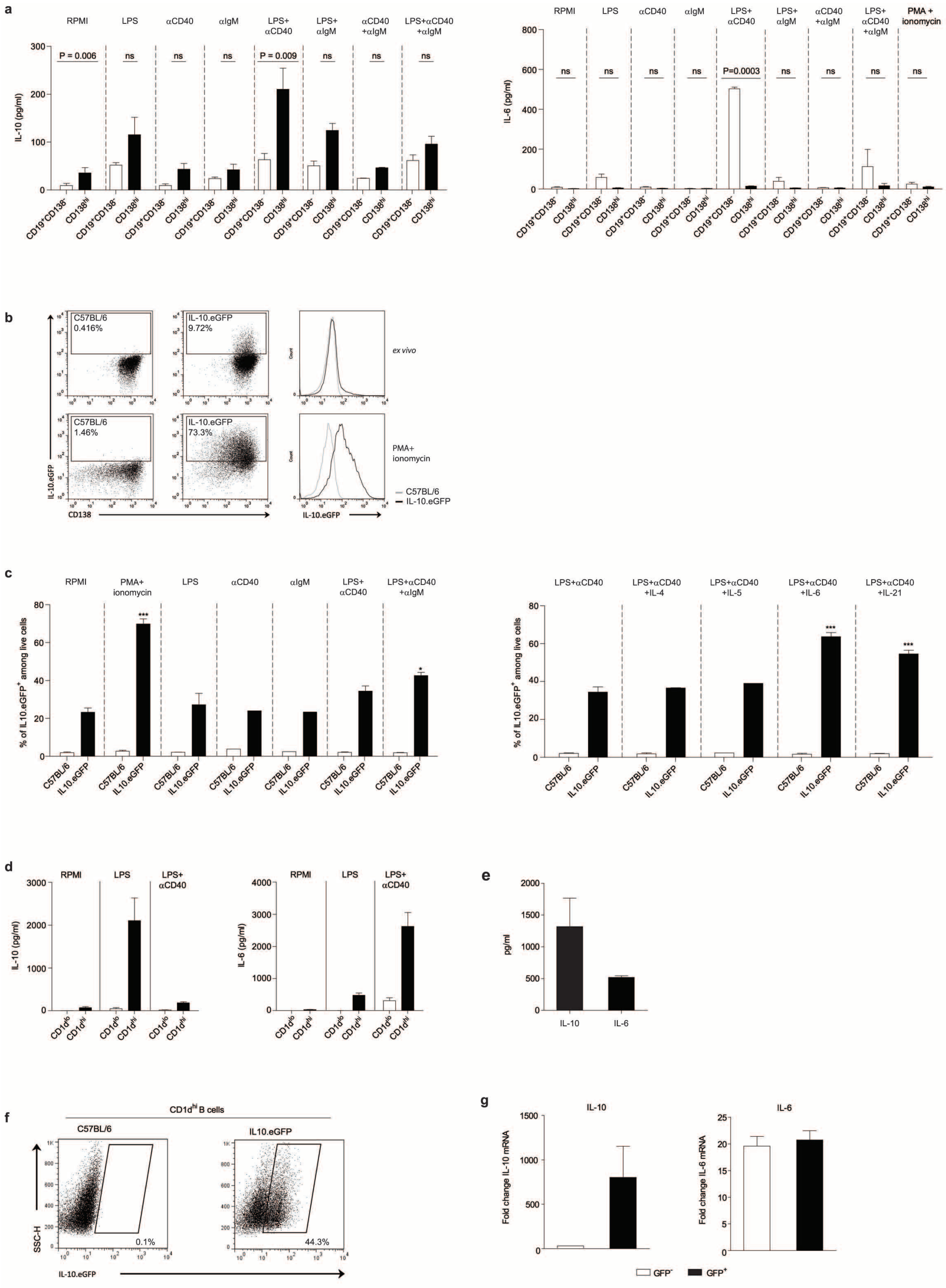
Extended Data Figure 5 | Role of IL-35 and CD40 expression by B cells during *Salmonella* infection. **a**, $B^{p35-/-}$, $B^{Ebi3-/-}$, $B^{p40-/-}$ and their corresponding B^{WT} mice were infected with attenuated *Salmonella* (strain SL7207). Relative titres of *Salmonella*-reactive IgM and IgG in serum were determined by ELISA. Data (mean ± s.e.m.) are pooled from three independent experiments. Numbers in brackets indicate the numbers of mice analysed for each measurement. **b**, $B^{p35-/-}$ ($n = 17$) and corresponding B^{WT} mice ($n = 16$) were immunized with 200 μg NP-OVA in alum intraperitoneally. Relative titres of NP(15)-BSA-reactive antibodies in serum were determined by ELISA. Data (mean ± s.e.m.) are pooled from three independent experiments. Results were compared using two-tailed unpaired *t*-test. *P* values >0.05 are indicated by ns. **c**, Mice were infected with 100 c.f.u. of virulent *Salmonella*. Upper left shows survival curves of $B^{Cd40-/-}$ ($n = 9$) and corresponding B^{WT} mice ($n = 16$). Data are pooled from two independent experiments. Survival curves were compared using Wilcoxon test. Upper right, numbers of inflammatory foci per liver section were determined on day 5 p.i. Lower panel, bacterial loads in liver (left) and spleen (right) were determined on day 5 p.i. Data (mean ± s.e.m.) are pooled from two independent experiments. Numbers of mice analysed: $B^{Cd40-/-}$ ($n = 12$), and corresponding B^{WT} mice ($n = 13$). Results were analysed using two-tailed unpaired *t*-test with Welch's correction.



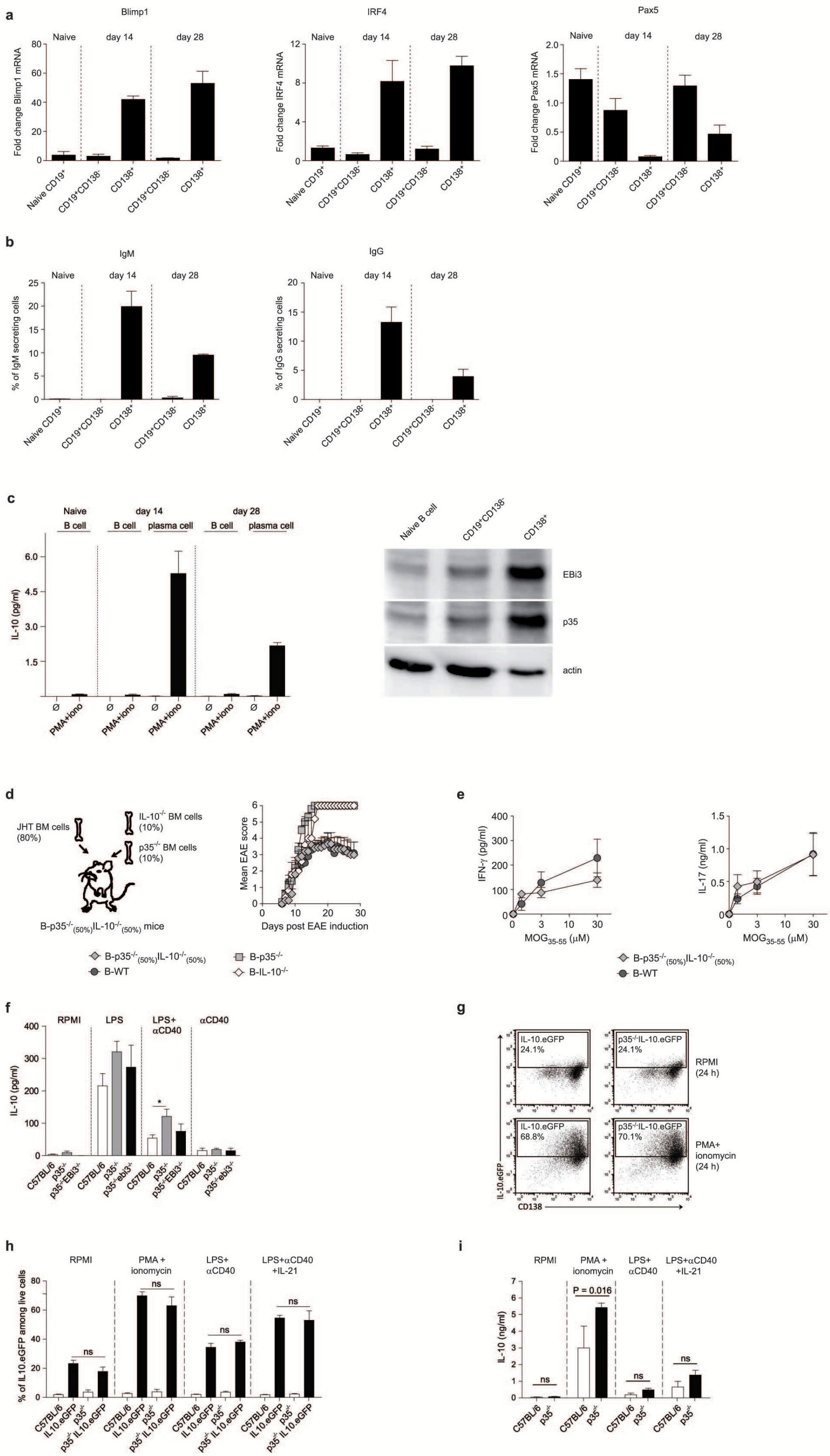
Extended Data Figure 6 | Characterization of IL-10- and IL-35-expressing plasma cells during *Salmonella* infection. **a**, **b**, C57BL/6 mice were infected intravenously with attenuated *Salmonella* (strain SL7207; 10^7 c.f.u.). On indicated days p.i., $CD138^{hi}$ plasma cells and $CD19^+CD138^-$ B cells were isolated from spleen through magnetic and FACS procedures. **a**, Strategy for isolation of $CD138^{hi}$ plasma cells and $CD19^+CD138^-$ B cells: splenocytes were stained with anti-CD138-PE followed by labelling with magnetic anti-PE microbeads, and magnetic separation on autoMACS (Miltenyi Biotech). The obtained CD138-positive and CD138-negative fractions were stained for CD19, CD138 and CD11b/CD11c/TCR-β/DAPI. $CD138^{hi}CD11b^-CD11c^-TCR\text{-}\beta^-DAPI^-$ and $CD19^+CD138^-CD11b^-CD11c^-TCR\text{-}\beta^-DAPI^-$ cells were then isolated from CD138-positive and CD138-negative fraction, respectively, by FACS, as bulk or single cells. Numbers in FACS plots indicate the percentages of $CD19^+CD138^-$ B cells and $CD138^{hi}$ plasma cells in total splenocytes on day 3 after challenge (left plots), after magnetic isolation (middle plots), and after FACS isolation (right plots). **b**, Single $CD138^{hi}$ cells were isolated on indicated days, and analysed by single cell-PCR using a mix of primers for β-actin, *Il10*, *p35* and *Ebi3*. Only cells giving a positive signal for β-actin were included in the calculations shown. Data show frequencies of $CD138^{hi}$ cells that gave a positive signal for *Il10*, both *p35* and *Ebi3*, or *Il10* and *p35* and *Ebi3*. Numbers of cells tested: day 1 (181 cells; 86% positive for β-actin), day 3 (156 cells; 79% positive for β-actin) and day 8 (163 cells; 97% positive for β-actin). Cells were obtained from two independent experiments. **c**, **d**, $CD138^{hi}$ plasma cells and $CD19^+CD138^-$ B cells were enriched by magnetic isolation (as described in **a**) from spleen of C57BL/6 mice on day 3 after infection with attenuated *Salmonella* (SL7207; 10^7 c.f.u.). **c**, Data show expression levels of selected surface receptors on live ($DAPI^-$) $CD138^{hi}$ plasma cells and $CD19^+CD138^-$ B cells. Cells were gated as in **a** (middle panels), and analysed by flow cytometry. **d**, Levels of CD1d and Tim-1 expressed by $CD138^{hi}$ cells were compared to those found respectively on $CD1d^{hi}$ and $CD1d^{lo}$ B cells, and on $Tim1^+$ and $Tim1^-$ B cells from spleen of naive C57BL/6 mice. Data show representative results from 2–5 independent experiments.


a
day 3
splenocytes
MACS
CD138
negative
fraction
CD138
positive
fraction
P1
CD19
CD138
P4
P3
P2
CD138
CD22
CD19+CD138- cells (P1)
CD138int CD22+ cells (P2)
CD138hi CD22+ cells (P3)
CD138hi CD22- cells (P4)
b
IgM
% of IgM secreting cells
IgG
% of IgG secreting cells
CD19+CD138-
CD138intCD22+
CD138hiCD22+
CD138hiCD22-
c
Blimp1
Fold change Blimp1 mRNA
IRF4
Fold change IRF4 mRNA
Pax5
Fold change Pax5 mRNA
IL-10
Fold change IL-10 mRNA
EBi3
Fold change EBi3 mRNA
p35
Fold change p35 mRNA
Naïve CD19+
CD19+CD138-
CD138intCD22+
CD138hiCD22+
CD138hiCD22-
d
IgKappa IgLambda / CD90 / CD22
Igkappa IgLambda / CD11b / CD22
IgKappa IgLambda / CD169 / CD22
100 µm

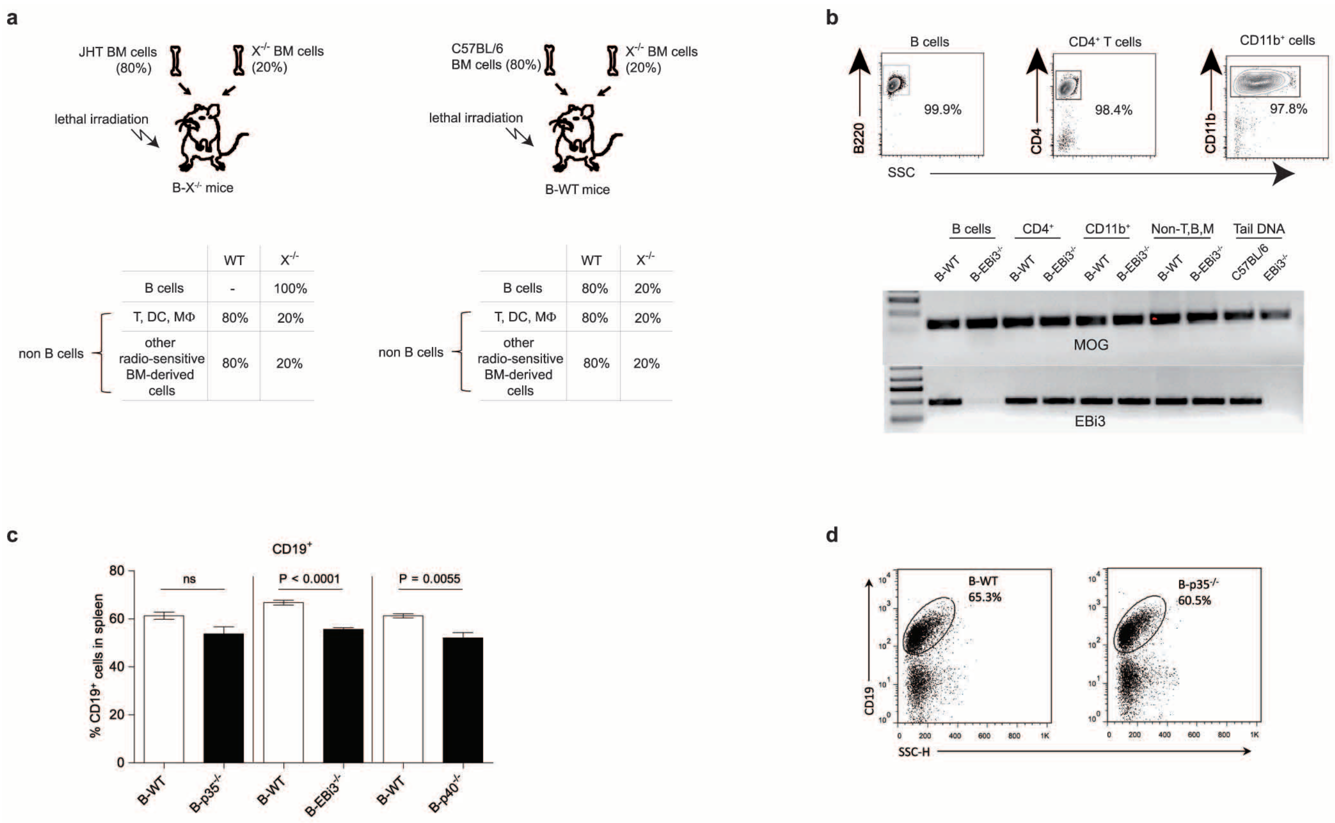
Extended Data Figure 7 | Characterization of plasma cells during *Salmonella* infection. **a**, $CD138^{hi}$ plasma cells and $CD19^{+}CD138^{-}$ B cells were magnetically enriched from spleens of C57BL/6 mice on day 3 after infection with attenuated *Salmonella* (SL7207; 10^{7} c.f.u.). $CD19^{+}CD138^{-}$ B cells (P1) were then isolated from the CD138-negative fraction by FACS. $CD138^{int}CD22^{+}$ (P2), $CD138^{hi}CD22^{+}$ (P3), and $CD138^{hi}CD22^{-}$ (P4) were isolated from the CD138-positive fraction by FACS. **b**, Frequencies of IgM- and IgG-secreting cells among B cells ($CD19^{+}CD138^{-}$) and each of these plasma cell subsets ($CD138^{int}CD22^{+}$, $CD138^{hi}CD22^{+}$, $CD138^{hi}CD22^{-}$) were determined by ELISPOT assay. A compilation of two independent experiments is shown (mean ± s.e.m.). **c**, *Blimp1*, *Irf4*, *Pax5*, *Il10*, *Ebi3* and *p35* mRNA expression was quantified by real-time PCR in these populations. Naive splenic B cells (naive $CD19^{+}$) were isolated from unchallenged C57BL/6 mice by magnetic selection. Data show the compilation of three independent experiments (mean ± s.e.m.). **d**, Plasma cells accumulate in splenic red pulp aggregates adjacent to white pulp areas on day 3 after *Salmonella* infection. C57BL/6 mice were infected intravenously with attenuated *Salmonella* (SL7207; 10^{7} c.f.u.), and spleens were harvested 3 days later. Spleen sections (7 μm) were stained with: top panels: anti-Ig-κ (clone 187.1; green), anti-Igλ (clone SL136; green), anti-CD90 (clone T24; red) and anti-CD22 (clone OX-97; blue); middle panels: anti-Ig-κ (clone 187.1; red), anti-Igλ (clone SL136; red), anti-CD11b (clone M1/70; green), and anti-CD22 (clone OX-97; blue); bottom panels: anti-Ig-κ (clone 187.1; red), anti-Igλ (clone SL136; red), anti-CD169 (clone MOMA-1; green), and anti-CD22 (clone OX-97; blue).


a
RPMI
LPS
αCD40
αIgM
LPS+ αCD40
LPS+ αIgM
αCD40 +αIgM
LPS+αCD40 +αIgM
P = 0.006
ns
P = 0.009
IL-10 (pg/ml)
CD19+CD138-
CD138hi
PMA + ionomycin
P=0.0003
IL-6 (pg/ml)
b
C57BL/6 0.416%
IL-10.eGFP 9.72%
ex vivo
C57BL/6 1.46%
IL-10.eGFP 73.3%
PMA+ ionomycin
C57BL/6
IL-10.eGFP
CD138
Count
c
% of IL10.eGFP+ among live cells
LPS+αCD40 +IL-4
LPS+αCD40 +IL-5
LPS+αCD40 +IL-6
LPS+αCD40 +IL-21
d
CD1dlo
CD1dhi
e
pg/ml
IL-10
IL-6
f
CD1dhi B cells
IL10.eGFP
SSC-H
0.1%
44.3%
g
Fold change IL-10 mRNA
Fold change IL-6 mRNA
GFP-
GFP+

Extended Data Figure 8 | Production of IL-10 and IL-6 by plasma cells and B cells after *Salmonella* infection. **a**, CD138^{hi} plasma cells and CD19^{+}CD138^{-} B cells were isolated from spleen of C57BL/6 mice on day 3 after infection with attenuated *Salmonella* (SL7207; 10^7 c.f.u.), as described in Extended Data Fig. 6. Isolated cells were activated for 24 h as indicated with LPS (1 μg ml^{-1}), anti-CD40 (10 μg ml^{-1}), and anti-IgM (5 μg ml^{-1}). IL-10 (left) and IL-6 (right) concentrations were determined by Bio-Plex. Data shown are pooled from 2–5 independent experiments (mean ± s.e.m.). Results were compared using two-tailed unpaired *t*-test. P values > 0.05 are indicated by ns. **b**, **c**, CD138^{hi} plasma cells were isolated from spleen of C57BL/6 or IL-10–eGFP mice on day 3 after infection with attenuated *Salmonella* (SL7207; 10^7 c.f.u.). **b**, Cells were analysed by flow cytometry directly after isolation to quantify IL-10–eGFP expression (*ex vivo*), and 24 h after re-stimulation with PMA plus ionomycin. Data show representative FACS plots and histogram overlays. Numbers in the plots indicate the percentages of IL-10–eGFP-positive cells among total live (PI^{-}) cells. **c**, Cells isolated from C57BL/6 and IL-10–eGFP mice were re-stimulated for 24 h, as indicated with LPS (1 μg ml^{-1}), anti-CD40 (10 μg ml^{-1}), anti-IgM (5 μg ml^{-1}), IL-4 (20 ng ml^{-1}), IL-5 (20 ng ml^{-1}), IL-6 (20 ng ml^{-1}) and IL-21 (20 ng ml^{-1}). Cells were analysed as in **b**. Data shown are pooled from 2–4 independent experiments (mean ± s.e.m.). Results were compared between condition RPMI and activated conditions (left) or between condition with added cytokine and condition LPS plus anti-CD40 (right) using one-way ANOVA followed by Bonferroni post-test (*$P < 0.05$, **$P < 0.01$; ***$P < 0.001$). **d**–**g**, Expression of IL-6 by IL-10-expressing CD1d^{hi} B cells. **d**, CD19^{+}CD1d^{hi} and CD19^{+}CD1d^{lo} cells were isolated from spleen of naive C57BL/6 mice according to CD19 and CD1d expression using magnetic and FACS procedures, and stimulated for 24 h, as indicated. IL-6 and IL-10 concentrations in culture supernatants were determined by Bio-Plex. **e**, **f**, CD1d^{hi} B cells were isolated from naive IL-10–eGFP mice and stimulated with LPS (2 μg ml^{-1}) for 48 h. IL-6 and IL-10 concentrations in culture supernatants were determined by Bio-Plex (**e**), and cells were analysed by flow cytometry to determine the frequency of IL-10–eGFP-expressing cells (**f**). **g**, CD1d^{hi} B cells were isolated from naive IL-10–eGFP mice, and stimulated with LPS (2 μg ml^{-1}) for 48 h. IL-10–eGFP^{-} and IL-10–eGFP^{+} cells were then separated by FACS and analysed by real time PCR for expression levels of *Il10* and *Il6* mRNA. Data shown are pooled from two independent experiments (mean ± s.e.m.).


a
Blimp1
IRF4
Pax5
Naive
day 14
day 28
Fold change Blimp1 mRNA
Fold change IRF4 mRNA
Fold change Pax5 mRNA
Naive CD19+
CD19+CD138-
CD138+
b
IgM
IgG
% of IgM secreting cells
% of IgG secreting cells
c
B cell
plasma cell
IL-10 (pg/ml)
Ø
PMA+iono
Naive B cell
EBi3
p35
actin
d
JHT BM cells (80%)
IL-10-/- BM cells (10%)
p35-/- BM cells (10%)
B-p35-/-(50%)IL-10-/-(50%) mice
Mean EAE score
Days post EAE induction
B-p35-/-(50%)IL-10-/-(50%)
B-WT
B-p35-/-
B-IL-10-/-
e
IFN-γ (pg/ml)
IL-17 (ng/ml)
MOG35-55 (μM)
f
RPMI
LPS
LPS+ αCD40
αCD40
IL-10 (pg/ml)
C57BL/6
p35-/-
p35-/-EBi3-/-
g
IL-10.eGFP 24.1%
p35-/-IL-10.eGFP 24.1%
RPMI (24 h)
IL-10.eGFP 68.8%
p35-/-IL-10.eGFP 70.1%
PMA+ ionomycin (24 h)
IL-10.eGFP
CD138
h
PMA + ionomycin
LPS+ αCD40
LPS+αCD40 +IL-21
% of IL10.eGFP among live cells
ns
IL10.eGFP
p35-/- IL10.eGFP
i
IL-10 (ng/ml)
P = 0.016

Extended Data Figure 9 | IL-35 and IL-10 production by plasma cells during EAE and *Salmonella* infection. **a**, B cells and plasma cells were isolated from spleens of mice on days 14 and 28 after EAE induction through magnetic and FACS procedures. Splenocytes were stained with anti-CD138-PE followed by labelling with anti-PE microbeads, and magnetic separation. The obtained CD138-positive and CD138-negative fraction were stained for CD19, CD138, and CD11b/CD11c/CD90/DAPI. $CD138^+CD11b^-CD11c^-CD90^-DAPI^-$ and $CD19^+CD138^-CD11b^-CD11c^-CD90^-DAPI^-$ cells were then isolated from CD138-positive and CD138-negative fractions, respectively, by FACS, as bulk cells. Naive splenic B cells (naive $CD19^+$) were isolated from unchallenged C57BL/6 mice by magnetic selection. *Blimp1*, *Irf4* and *Pax5* mRNA expression was quantified by real-time PCR. Data compile 2 independent experiments (mean ± s.e.m.). **b**, Production of IgM and IgG by plasma cells and B cells isolated from mice during EAE was determined by ELISPOT. A compilation of 3 independent experiments for day 14, and 2 independent experiments for day 28 is shown (mean ± s.e.m.). **c**, Left, $CD19^+CD138^-$ B cells (B cell) and $CD138^+$ plasma cells (plasma cell) were isolated from spleens of mice on days 14 and 28 after EAE induction through magnetic and FACS procedures, as described in **a**, and stimulated for 24 h with PMA plus ionomycin (PMA plus iono) or kept in medium (∅). IL-10 concentrations were determined by Bio-Plex. Data are pooled from 4 independent experiments for day 14 and 2 independent experiments for day 28 (mean ± s.e.m.). Right, B cells ($CD19^+CD138^-$) and $CD138^+$ plasma cells ($CD138^+$) were isolated from spleens of mice on day 14 after EAE induction through magnetic isolation. For this, splenocytes were stained with anti-CD138-PE followed by labelling with anti-PE microbeads, and magnetic separation to yield $CD138^+$ plasma cells. B cells were obtained from the CD138-negative fraction after further depletion of remaining $CD138^+$ cells, and positive selection using anti-CD19 magnetic microbeads. Purities were above 95%. Cell lysates were separated on SDS–PAGE gel and sequentially blotted with anti-EBI3, anti-p35, and anti-actin antibodies. Data show representative results of two independent experiments. **d**, **e**, Mice in which B cells cannot co-express IL-10 and IL-35 do not develop exacerbated EAE. **d** Left, $B^{p35^{-/-}}{}_{(50\%)}{}^{Il10^{-/-}}{}_{(50\%)}$ were obtained by reconstituting lethally irradiated recipient mice with BM cells from JHT mice (80%), IL-10-deficient mice (10%), and p35-deficient mice (10%). B^{WT} mice were produced using C57BL/6 BM instead of JHT BM cells. $B^{p35^{-/-}}$ and $B^{Il10^{-/-}}$ mice were obtained as outlined in Extended Data Fig. 10. Right, EAE was induced in $B^{p35^{-/-}}{}_{(50\%)}{}^{Il10^{-/-}}{}_{(50\%)}$ mice (grey diamonds, $n = 10$), $B^{p35^{-/-}}$ (grey squares, $n = 10$), $B^{Il10^{-/-}}$ (white diamonds, $n = 5$) and B^{WT} (dark grey circles, $n = 10$) mice. Data show clinical EAE scores. **e**, Splenocytes were harvested from $B^{p35^{-/-}}{}_{(50\%)}{}^{Il10^{-/-}}{}_{(50\%)}$ and B^{WT} mice on day 28 after EAE induction, and re-stimulated individually for 48 h with $MOG_{35–55}$. Culture supernatants were analysed by ELISA to determine concentrations of IFN-γ and IL-17. Data (mean ± s.e.m.) are pooled from two independent experiments. **f**, Production of IL-10 by IL-35-deficient B cells *in vitro*. Splenic B cells from naive C57BL/6, $p35^{-/-}$, and $p35^{-/-}Ebi3^{-/-}$ mice were stimulated with LPS (1 μg ml^{-1}), anti-CD40 (10 μg ml^{-1}), alone or in combination, as indicated, for 72 h, and IL-10 concentrations were determined by Bio-Plex. Data (mean ± s.e.m.) are pooled from three independent experiments. Results were compared between C57BL/6 cells and cells from mutant mice using one-way ANOVA followed by Bonferroni post-test ($*P < 0.05$). **g**–**i**, IL-35 is not necessary for production of IL-10 by $CD138^{hi}$ plasma cells during *Salmonella* infection. $CD138^{hi}$ plasma cells were isolated from spleen of C57BL/6, IL-10–eGFP, $p35^{-/-}$, and $p35^{-/-}$ IL-10–eGFP mice on day 3 after infection with attenuated *Salmonella* (10^7 c.f.u.) using a combination of magnetic and FACS methods. Cells were analysed by flow cytometry (**g**) to determine the frequency of IL-10–eGFP-expressing cells after re-stimulation as indicated for 24 h (**h**). Numbers in the plots indicate the percentages of IL-10–eGFP-positive cells among total live (PI^-) cells. Results were compared using two-tailed unpaired *t*-test. **i**, $CD138^{hi}$ plasma cells isolated from C57BL/6 and $p35^{-/-}$ mice were re-stimulated for 24 h as indicated. IL-10 concentrations were determined by Bio-Plex. Data are pooled from two independent experiments (mean ± s.e.m.). Results were compared using two-tailed unpaired *t*-test with Welch's correction. *P* values >0.05 are indicated by ns.



Extended Data Figure 10 | Generation and validation of bone marow chimaera. **a**, Scheme for the generation of $B^{X-/-}$ and corresponding control B^{WT} mice. Left, $B^{X-/-}$ mice with a deficiency in gene *X* restricted to B cells were obtained by reconstituting lethally irradiated recipient mice with a mixture of BM cells from B-cell-deficient JHT mice (80%) and from gene *X*-deficient ($X^{-/-}$) mice (20%). The table indicates the contribution of JHT and $X^{-/-}$ BM cells to relevant haematopoietic cell types in the resulting chimaera, taking into account that JHT mice carry wild-type (WT) alleles of gene *X*. In $B^{X-/-}$ mice, B cells only developed from $X^{-/-}$ BM cells and therefore were all $X^{-/-}$. In contrast, all other haematopoietic lineages developed from both $X^{-/-}$ and JHT BM cells. As the JHT and $X^{-/-}$ BM cells were injected in a ratio 80:20, most haematopoietic cells other than B cells carried a wild-type genotype for *X* in $B^{X-/-}$ mice. Right, the corresponding control B^{WT} mice were obtained by reconstituting lethally irradiated recipient mice with a mixture of BM cells from WT C57BL/6 mice (80%) and from $X^{-/-}$ mice (20%). The table indicates the contribution of WT and $X^{-/-}$ BM cells to relevant haematopoietic cell types in the resulting chimaera. In B^{WT} mice, all haematopoietic lineages, including B cells, developed from WT (80%) and $X^{-/-}$ (20%) BM cells. Therefore B cells mostly had a wild-type genotype in B^{WT} mice, which also controlled for possible effects owing to the 20% $X^{-/-}$ haematopoietic cells other than B cells present in $B^{X-/-}$ mice (for example, 20% $X^{-/-}$ T cells, and so on). **b**, Top, representative FACS plots of isolated B cells, $CD4^+$ T cells and $CD11b^+$ cells purified from naive $B^{Ebi3-/-}$ and B^{WT} mice, using magnetic and FACS-based methodologies; (bottom) genomic DNA was extracted from the sorted cell populations, including the non-labelled $B220^-CD4^-CD11b^-$ cells (indicated as Non-T,B,M). Genomic DNA samples from tail biopsies of C57BL/6 and *Ebi3*$^{-/-}$ mice were used as controls. The amounts of genomic DNA present in each sample were equilibrated using the *Mog* gene as standard. Using equilibrated samples, the wild-type *Ebi3* allele was detected by PCR in all DNA preparations except for B cells from $B^{Ebi3-/-}$ mice, and the tail biopsy from *Ebi3*$^{-/-}$ mice. **c**, Reconstitution of the B-cell compartment in BM chimaeric mice. $B^{p35-/-}$, $B^{Ebi3-/-}$, $B^{p40-/-}$ and corresponding B^{WT} controls were killed after reconstitution. Data show the percentage of $CD19^+$ B cells in spleen, as determined by flow cytometry. Data are pooled from 2–3 independent experiments. Data (mean ± s.e.m.) include $B^{p35-/-}$ ($n = 6$) and corresponding B^{WT} mice ($n = 7$), $B^{Ebi3-/-}$ ($n = 7$) and corresponding B^{WT} mice ($n = 9$), $B^{p40-/-}$ ($n = 7$) and corresponding B^{WT} mice ($n = 7$). Comparisons of B-cell reconstitutions between $B^{X-/-}$ and corresponding B^{WT} mice were done with two-tailed unpaired *t*-test with Welch's correction (*P* values > 0.05 are indicated as ns). **d**, Representative FACS plots of B cells from B^{WT} and $B^{p35-/-}$ mice. Cells are gated on live (PI^-) spleen cells.

LETTER

doi:10.1038/nature13138

Obesity-associated variants within *FTO* form long-range functional connections with *IRX3*

Scott Smemo[1]*, Juan J. Tena[2]*, Kyoung-Han Kim[3]*, Eric R. Gamazon[4], Noboru J. Sakabe[1], Carlos Gómez-Marín[2], Ivy Aneas[1], Flavia L. Credidio[1], Débora R. Sobreira[1], Nora F. Wasserman[1], Ju Hee Lee[3], Vijitha Puviindran[3], Davis Tam[3], Michael Shen[1], Joe Eun Son[5], Niki Alizadeh Vakili[3], Hoon-Ki Sung[5], Silvia Naranjo[2], Rafael D. Acemel[2], Miguel Manzanares[6], Andras Nagy[5], Nancy J. Cox[1,4], Chi-Chung Hui[3], Jose Luis Gomez-Skarmeta[2] & Marcelo A. Nóbrega[1]

Genome-wide association studies (GWAS) have reproducibly associated variants within introns of *FTO* with increased risk for obesity and type 2 diabetes (T2D)[1–3]. Although the molecular mechanisms linking these noncoding variants with obesity are not immediately obvious, subsequent studies in mice demonstrated that *FTO* expression levels influence body mass and composition phenotypes[4–6]. However, no direct connection between the obesity-associated variants and *FTO* expression or function has been made[7–9]. Here we show that the obesity-associated noncoding sequences within *FTO* are functionally connected, at megabase distances, with the homeobox gene *IRX3*. The obesity-associated *FTO* region directly interacts with the promoters of *IRX3* as well as *FTO* in the human, mouse and zebrafish genomes. Furthermore, long-range enhancers within this region recapitulate aspects of *IRX3* expression, suggesting that the obesity-associated interval belongs to the regulatory landscape of *IRX3*. Consistent with this, obesity-associated single nucleotide polymorphisms are associated with expression of *IRX3*, but not *FTO*, in human brains. A direct link between *IRX3* expression and regulation of body mass and composition is demonstrated by a reduction in body weight of 25 to 30% in *Irx3*-deficient mice, primarily through the loss of fat mass and increase in basal metabolic rate with browning of white adipose tissue. Finally, hypothalamic expression of a dominant-negative form of *Irx3* reproduces the metabolic phenotypes of *Irx3*-deficient mice. Our data suggest that *IRX3* is a functional long-range target of obesity-associated variants within *FTO* and represents a novel determinant of body mass and composition.

Noncoding variation in single nucleotide polymorphisms (SNPs) within a 47-kilobase (kb) region of high linkage disequilibrium in introns 1 and 2 of *FTO* remains the strongest genetic association with risk to polygenic obesity in humans[1–4]. Individuals homozygous for risk alleles of the associated SNPs weigh approximately 3 kg more than individuals homozygous for non-risk alleles, underscoring the significant phenotypic impact of these common variants[2]. Multiple follow-up studies directly implicated *FTO* as a gene controlling body mass and composition. Null *Fto* alleles in mice result in lean phenotypes, with reciprocal models overexpressing *FTO* displaying increased body weight[4–6]. However, there is no direct evidence that enhancers within the obesity-associated *FTO* introns are connected with regulation of *FTO* expression. Expression quantitative trait locus (eQTL) analyses have systematically failed to show association between the obesity-associated SNPs and *FTO* expression in human tissues[7–9].

To chart directly the *cis*-regulatory circuitry within the *FTO* locus, we used circular chromosome conformation capture followed by high-throughput sequencing (4C-seq). We carried out 4C-seq in whole mouse embryos (embryonic day 9.5; E9.5) and in adult (8 weeks) mouse brains, as previous work suggests that brain *FTO* expression modulates metabolic parameters[6,10]. We profiled the genomic interactions with promoters of genes located within a 1-megabase (Mb) window around the obesity-associated SNPs (mouse genome build mm9, chromosome 8: 93,725,000–94,725,000), including *Fto* and *Rpgrip1l*, and *Irx3*, a half-megabase downstream. There was a clear difference in interaction patterns among genes (Fig. 1a). The *Fto* promoter chiefly participates in genomic interactions proximal to the gene promoter. Although in mouse embryos these interactions include the obesity-associated intronic region (Fig. 1b), no such interactions are detected in adult mouse brains (Extended Data Fig. 1 and Supplementary Table 1). In marked contrast, the promoter of *Irx3* participates in numerous long-range interactions across a broad genomic region encompassing nearly 2 Mb, including robust interactions with the obesity-associated interval within *FTO* (Fig. 1a). We confirmed these interactions between *Irx3* and the *Fto* obesity-associated region using chromatin conformation capture (3C) in adult mouse brains (Extended Data Fig. 2c). These data suggest that the obesity-associated interval is likely to be part of the regulatory landscape of *Irx3*. We next inferred that the long-range interactions between the obesity-associated *FTO* intron and *IRX3* represent a conserved feature in vertebrate genomes. Inspecting the human ENCODE data set, we observe that human breast cancer MCF-7 cells display the same pattern of long-range looping between the obesity-associated interval and *IRX3*, but not *FTO*, assayed by chromatin interaction analysis with paired-end tag sequencing (ChIA-PET) (Extended Data Fig. 2a). This is corroborated further by recent Hi-C (high-throughput 3C) data in human fibroblasts[11] (Extended Data Fig. 2b). We also observed a similar pattern of chromatin interactions in zebrafish embryos, assayed by 4C-seq (Extended Data Fig. 1c). Together, these data suggest that the obesity-associated *FTO* intron is mediating functional interactions with *IRX3* in the human, mouse and zebrafish genomes.

After revealing evidence of direct looping between *FTO* intronic regions and *IRX3*, we next showed that enhancers within the obesity-associated interval possess functional characteristics compatible with *IRX3* expression. We observed that the degree of evolutionary conservation in noncoding sequences of the *FTO-IRX* locus (chromosome 16: 53,731,249–54,975,288) places it on the top 2% of similarly sized genomic intervals, suggesting the presence of many functional noncoding elements spread over the region (see Methods). More specifically, ENCODE data suggest that the 47-kb obesity-associated interval is fraught with *cis*-regulatory elements, evidenced by an abundance of enhancer-associated chromatin marks, DNase hypersensitive sites, and transcription-factor binding events across this genomic region (Fig. 1b). We tested three human DNA fragments from the 47-kb obesity-associated region for their putative enhancer properties using an *in vivo* mouse reporter assay. These fragments overlap enhancer-like chromatin marks in multiple cell lines. All three fragments displayed enhancer activity in neonatal (postnatal day

[1]Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA. [2]Centro Andaluz de Biología del Desarrollo (CABD), Consejo Superior de Investigaciones Científicas/Universidad Pablo de Olavide, Carretera de Utrera Km1, Sevilla 41013, Spain. [3]Program in Developmental & Stem Cell Biology, The Hospital for Sick Children, and Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [4]Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, Illinois 60637, USA. [5]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario M5T 3H7, Canada. [6]Cardiovascular Development and Repair Department, Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid 28029, Spain.
*These authors contributed equally to this work.

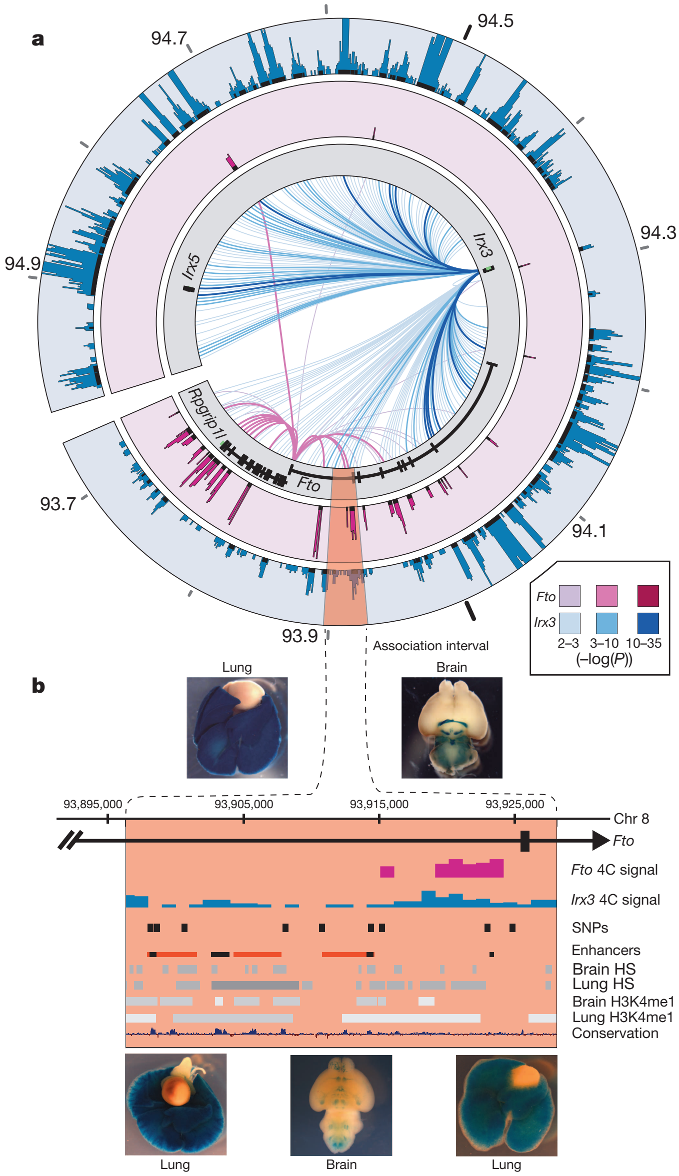


Figure 1 | Long-range interactions in the *IRX3-FTO* locus. a, Mouse embryo 4C-seq interactions emanating from each promoter are displayed as links across the circle (darker link implies greater significance). Outer plots show significance of interactions above background (−log(*P* value)). The obesity-associated interval is highlighted red. **b**, Magnified view of the association interval. Contained are the orthologous locations of obesity-associated SNPs (black pips), and epigenetic marks associated with regulatory elements. Endogenous *Irx3* expression is shown in lung and brain in *Irx3*^{lacZ} knock-in mouse (top two images). Three enhancers drive reporter expression in lungs and brain (rectangles in the Enhancers row, corresponding to bottom three images). Other enhancers[17] are shown in black. HS, DNase hypersensitive sites.

0 (P0)–P1) mice. Interestingly, two of the fragments drove strong reporter gene expression in lungs (Fig. 1b). *Irx3* is highly expressed during lung development (Fig. 1b)[12,13], whereas *Fto* expression in lungs is limited (Extended Data Fig. 3a and refs 2, 14–16). Consistent with our data, a previous report has shown that the *FTO* obesity-associated region contains enhancers that drive *Irx3*-like expression in zebrafish[17].

Our data suggest that the regulatory landscape of *IRX3* spreads over megabase distances, whereas *FTO* expression is primarily regulated by regions proximal to its promoter. To test this, we engineered a human bacterial artificial chromosome (BAC) spanning 162 kb of the *FTO* locus, including its promoter and the 47-kb obesity-associated region (Extended Data Fig. 3). We recombineered a reporter cassette at the *FTO* translation start, and generated transgenic mice harbouring the engineered BAC. Transgenic mice expressed the reporter gene in multiple tissues, recapitulating the endogenous expression pattern of *FTO*[14]. We next determined that a 1.2-kb region corresponding to the *FTO* promoter is sufficient to recapitulate most of the *FTO* expression pattern (Extended Data Fig. 3). In contrast, a 2.8-kb region corresponding to the *IRX3* promoter does not recapitulate any of the endogenous expression patterns of *IRX3*, suggesting that *IRX3* expression relies on long-range *cis*-regulatory elements. A total of 15 human sequences in the broader *FTO-IRX* locus have been characterized as *in vivo* enhancers in mouse[17,18], driving reporter expression in *Irx3*-expressing tissues, including eye, limb, brain, neural tube and branchial arches[12,19] (Supplementary Table 2). Although these enhancers lie outside the obesity association interval, they are consistent with our chromatin looping data pointing to a broad *Irx3* regulatory landscape, extending into *Fto*. Further supporting these findings, the genomic interval spanning *FTO* and *IRX3* has been proposed to be part of a single genomic regulatory block, based on the presence of extended synteny across deep phylogenies and a high density of highly conserved noncoding elements[17], as well as patterns of CTCF binding and chromatin interactions[20]. Under these models, our data support the idea that the broad expression patterns of *FTO* are primarily regulated by elements proximal to the promoter, and that *IRX3* is endowed with an ancient, extensive *cis*-regulatory circuitry extending into *FTO*.

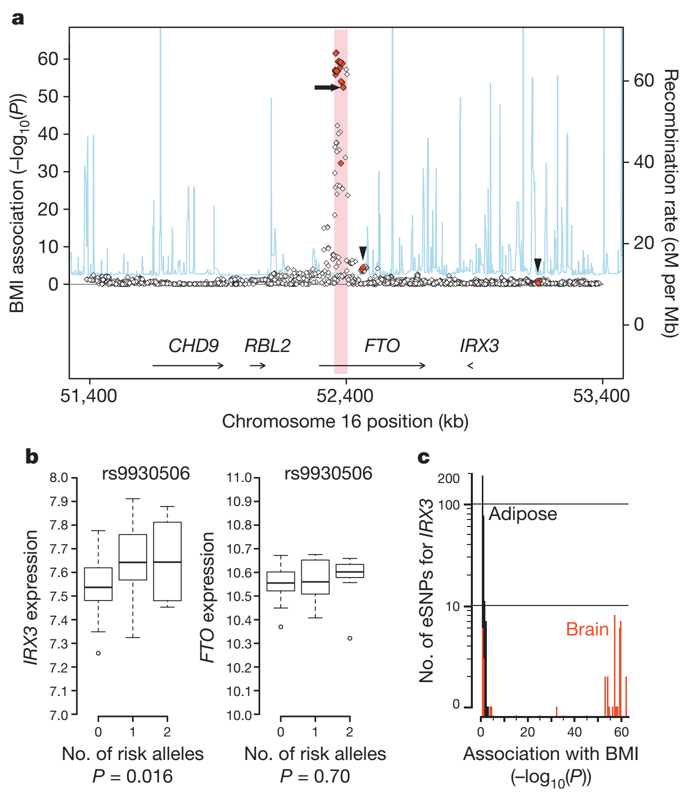


Figure 2 | BMI-associated SNPs are associated with expression of *IRX3*, but not *FTO*, in human brain. a, SNPs associated with BMI in the *FTO* locus. Forty-three SNPs associated with *IRX3* expression (eSNPs) are shown in red, including 11 also associated with BMI. No SNPs are associated with *FTO* expression ($P > 0.05$). Thick arrowheads, *IRX3* eSNPs outside the obesity-associated region (pink); thick arrow, rs9930506. **b**, In cerebellum, the allele of rs9930506 associated with increased BMI (risk allele) is correlated with increased *IRX3* expression and not with *FTO* expression. Vertical middle bars represent the first and third quartiles, horizontal lines represent the median, and open circles denote outliers beyond 1.5 times the interquartile range. **c**, Histogram plotting the number of SNPs associated with *IRX3* expression in adipose (black) and brain (red) (*y* axis) by the significance of their association with BMI (*x* axis). Full details of statistics are provided in the Methods.

We next determined that gene expression studies in human brains corroborate our chromatin looping data, showing that the obesity-associated SNPs are associated with expression levels of *IRX3*, but not *FTO*. Our data and that of others[17] (Fig. 1b) demonstrate that the obesity-associated SNPs overlap several *in vivo* enhancers, raising the possibility that allelic variants may disrupt enhancer activity and alter expression of their target gene(s). To test this, we carried out eQTL mapping in human brain samples. *FTO* and *IRX3* are highly expressed in multiple regions of the brain including cerebellum and hypothalamus (Extended Data Fig. 4). Using a data set of 153 brain samples from individuals of European ancestry[21], represented by cerebellum, we found significant association between 11 SNPs previously associated with increased body mass index (BMI), and expression of *IRX3*, but not *FTO* (Fig. 2a and Supplementary Table 3). These common SNPs have minor allele frequencies ranging from 42.5 to 47.5% and are in strong linkage disequilibrium with each other (Extended Data Fig. 5). In all cases, such as for SNP rs9930506 (Fig. 2b), the allele associated with increased BMI is also associated with increased *IRX3* expression. Further analysis, using the GIANT Consortium data set of SNPs associated with BMI in 249,796 individuals[22], reveals that among SNPs significantly associated with *IRX3* expression in brain or adipose tissue, only those associated with *IRX3* in brain show highly significant associations with BMI (Fig. 2c). Taken together, our data directly tie the noncoding genetic variation within *FTO* to tissue-specific modulation of *IRX3*, but not *FTO*, expression in human brain.

Next, we exploited animal models to determine a potential role for *IRX3* expression in the regulation of BMI and/or metabolism. Mice homozygous for an *Irx3*-null allele (*Irx3*-knockout mice) are viable and fertile, with no evidence of embryonic lethality. We observed a 25–30% reduction of body weight in *Irx3*-knockout mice compared to control littermates (wild type), independent of gender (Fig. 3a and Extended Data Figs 6a, b and 7a, b). This difference becomes more pronounced if animals were subjected to a high-fat diet (HFD), with *Irx3*-knockout animals showing no significant body weight gain, contrasting to a 63% increase in control animals (Fig. 3a, b). Importantly, the percentage of fat mass in *Irx3*-knockout mice was significantly reduced without marked change of the lean mass ratio (Fig. 3c, d and Extended Data Fig. 7c, d). *Irx3*-knockout mice exhibit marked reduction in adiposity with smaller fat depots as well as reduced adipocyte size (Fig. 3e and Extended Data Figs 6b–d and 7e, f). These results were confirmed by differential gene expression of adiposity markers (that is, leptin, adiponectin and *mcp1*) in the perigonadal white adipose tissue (PWAT) of *Irx3*-knockout mice (Extended Data Fig. 6e). Importantly, *Fto* expression was not altered in the hypothalamus or PWAT of *Irx3*-knockout mice, suggesting that the lean phenotype of *Irx3*-knockout mice is not associated with *Fto* (Fig. 3f). As *Irx3*-knockout mice were resistant to HFD-induced obesity and metabolic disorder, such as hepatosteatosis (Fig. 3e), we examined glucose homeostasis by performing glucose tolerance tests (GTT) and insulin tolerance test (ITT) at different time points in the presence or absence of HFD. In 8-week-old mice, no difference in GTT was found (Fig. 3g). Although ageing and HFD led to glucose intolerance and insulin resistance in wild-type mice, *Irx3*-knockout mice showed none of these metabolic phenotypes (Fig. 3g–j).

Indirect calorimetric analysis showed higher energy expenditure in *Irx3*-knockout mice (Fig. 3k and Extended Data Fig. 8a, b, e). Importantly,

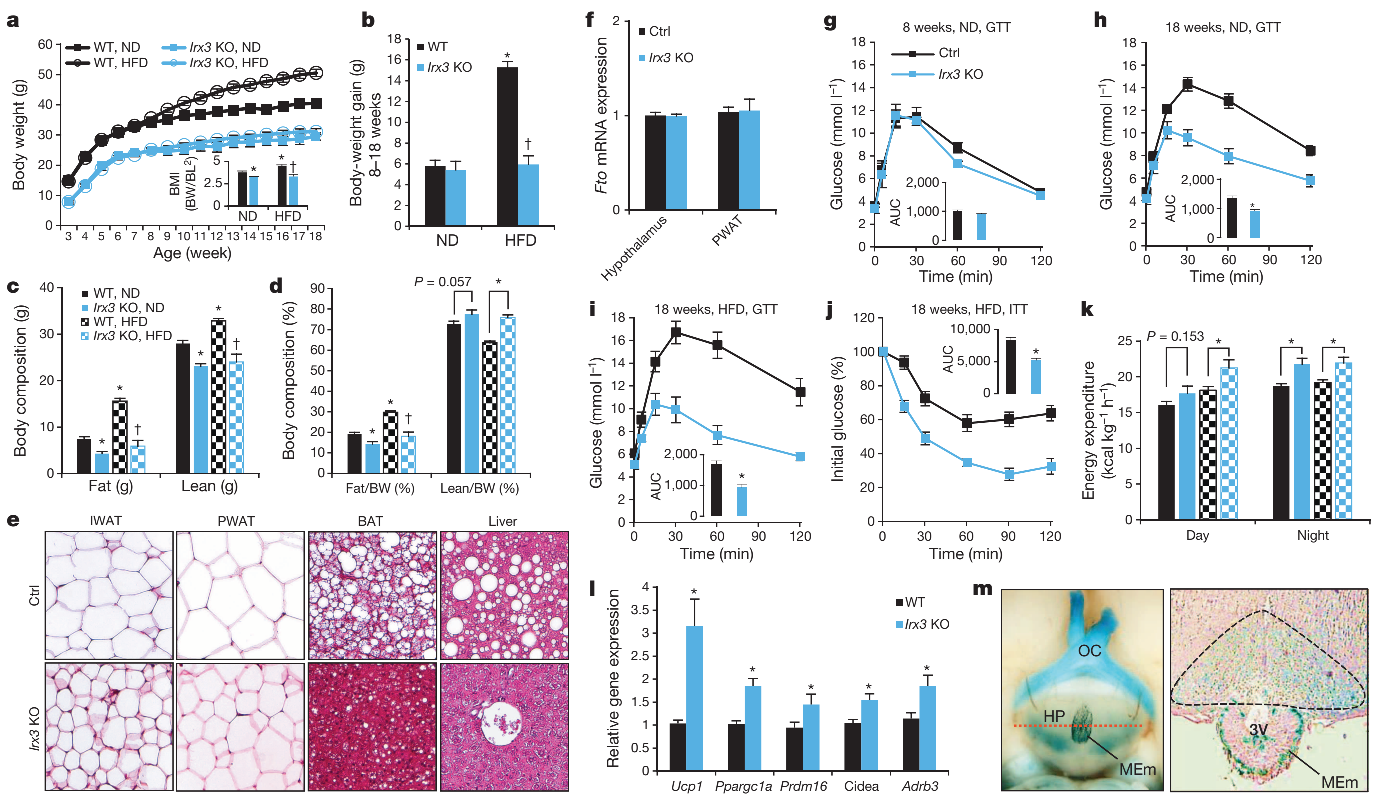


Figure 3 | *Irx3*-deficient mice are leaner and are protected against diet-induced obesity. **a**, Body weight in wild-type (WT) and *Irx3*-knockout (*Irx3* KO) mice fed a normal diet (ND) or high-fat diet (HFD). BL, body length. **b**, Weight gain in ND or HFD. **c**, Fat and lean mass in WT and *Irx3* KO mice. **d**, Fat and lean mass ratio as a percentage of body weight. **e**, Sections of inguinal white adipose tissue (IWAT; subcutaneous), perigonadal WAT (PWAT; visceral), brown adipose tissue (BAT), and liver from HFD mice. **f**, *Fto* messenger RNA expression in hypothalamus and PWAT of *Irx3* KO and WT mice. **g–i**, Glucose tolerance tests (GTT) in WT and *Irx3* KO mice. Inset graphs show area under curve (AUC). **j**, Insulin tolerance test (ITT) in HFD mice. **k**, Energy expenditure of ND and HFD mice. **l**, Gene expression in PWAT. **m**, *Irx3* expression in the arcuate nucleus and median eminence ($Irx3^{LacZ/+}$). Left panel shows a ventral view of a whole-mount stained brain. A dashed red line indicates the position of cross-section, displayed in the right panel. β-galactosidase-stained area in arcuate nucleus is marked by a dashed black line. 3V, third ventricle; HP, hypothalamus; MEm, median eminence; OC, optic chiasm. Data are expressed as mean ± s.e.m. $^*P < 0.05$ compared to WT or control; $^\dagger P < 0.05$ compared to each corresponding control group (see Methods). See additional statistical detail in Methods.

Irx3-knockout mice show upregulation of brown adipocyte markers in PWAT, including *Ucp1*, *Ppargc1a*, *Prdm16* and cidea, as well as increased expression of *Adrb3* encoding the β3-adrenergic receptor, suggestive of elevated sympathetic activation (Fig. 3l). Furthermore, we found a significant increase of *Ucp1* expression in the brown adipose tissue (BAT) of *Irx3*-knockout mice (Extended Data Fig. 8e, f). WAT predominantly stores excessive calories and its accumulation causes obesity, whereas brown adipose tissue dissipates energy (lipids and glucose) and has been the focus for potential development of novel therapeutic strategies to treat obesity and diabetes. Our data suggest that 'browning' of WAT by higher sympathetic tone and activation of BAT might lead to increased energy expenditure in *Irx3*-knockout mice and partially account for their leanness and protection from diet-induced obesity.

Browning of WAT by increased sympathetic activity is a phenomenon controlled by hypothalamic circuits integrating central and peripheral metabolism regulation. Our human eQTL data suggest that the obesity-associated SNPs are associated with *IRX3* expression in brain. We determined that *Irx3* is expressed in the arcuate nucleus and median eminence of the hypothalamus, two critical regions involved in the regulation of energy homeostasis (Fig. 3m)[23,24]. To investigate a possible role of *Irx3* expression in the hypothalamus mediating the body composition and energy homeostasis phenotypes in *Irx3*-knockout mice, we used *Rosa26*$^{EnR\text{-}Irx3}$ conditional transgenic mice expressing a dominant negative form of *Irx3* (*EnR-Irx3*) by crossing with *Ins2-Cre* (Extended Data Fig. 9a, b)[23,25]. This design allows for disrupting *Irx3* function while preserving the architecture of genomic interactions between the *Irx3* promoter and long-range regulatory sequences including the *Fto* obesity-associated region. Interestingly, *EnR-Irx3;Ins2-Cre* mice phenocopy the physiological, histological, and molecular metabolism phenotypes of germline *Irx3*-knockout mice (Fig. 4a–d and Extended Data Fig. 9c, d). Energy expenditure was significantly higher in *EnR-Irx3;Ins2-Cre* mice with similar food intake and locomotor activity (Fig. 4e and Extended Data Fig. 10a–e). Notably, these mice also showed 'browning' of WAT as well as activation of BAT (Fig. 4f and Extended Data Fig. 10f, g). Together, our data support the notion that the hypothalamic expression of *Irx3* regulates energy homeostasis and body composition.

Our results do not dispute the idea emerging from animal models that *Fto* is also a regulator of body mass and composition. The functional link between *FTO* and body mass regulation originated from animal models, unequivocally demonstrating that variation in *FTO* expression results in variation in body mass phenotypes. But to assess the null expectation that genetic manipulations in mice may result in variation in these parameters, we searched among 7,556 targeted mouse gene knockout models for evidence of alterations in body size, mass and growth (see Methods). Nearly one-third (2,166; 29%) of gene knockouts in mice show alteration of these phenotypes, underscoring that animal models alone are not sufficient to definitively establish a functional relationship between a given gene and long-range noncoding variants (Supplementary Table 4). It is the aggregate of our data on chromatin looping, patterns of *in vivo* enhancer function, human eQTL mapping and mouse model data that collectively place *IRX3* as an important target gene of the genetic association with human obesity. These observations have direct implications for how to interpret the associations characterized for thousands of noncoding variants with human disease and related quantitative traits. We observe more than 150 genomic regulatory blocks, similar to that containing *FTO* and *IRX3*, overlapping with noncoding SNPs associated with human disease and complex traits in GWAS (Methods and Supplementary Table 5), highlighting the need for a careful experimental pipeline to define the target gene(s) for each disease-associated noncoding SNP.

Our data posit that the obesity-associated SNPs within *FTO* are functionally connected with regulation of *IRX3* expression, and that *IRX3* is an important determinant of body mass and composition. *IRX3* encodes a transcription factor highly expressed in brain. Although our data represent the first demonstration of the intersection of *IRX3* biology with body mass composition and metabolism, previous work identified *IRX3* overexpression in adipocytes as a hallmark of the molecular switch seen in patients after profound weight loss following bariatric surgery[26]. These data indicate that IRX3 may have important roles regulating metabolism beyond the ones we describe associated with *Irx3* expression in hypothalamus. Future investigations will determine the precise molecular mechanisms by which *IRX3* regulates metabolic parameters.

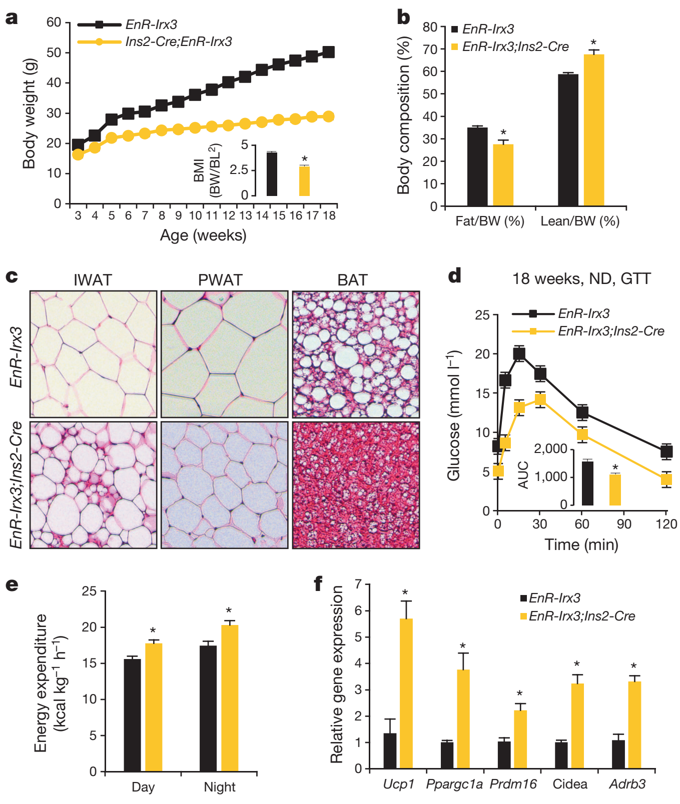


Figure 4 | Hypothalamus-specific dominant-negative *Irx3* mice recapitulate the metabolic phenotype of *Irx3*-deficient mice. **a**, Body weight of control (*EnR-Irx3*) and mutant mice (*EnR-Irx3;Ins2-Cre*) fed a normal diet. **b**, Fat and lean mass ratio as a percentage of body weight. **c**, Haematoxylin & eosin (H&E)-stained sections of IWAT, PWAT and BAT. **d**, Glucose tolerance test. **e**, Energy expenditure corrected for lean mass. **f**, Quantitative PCR in PWAT in control and mutant mice. Data are expressed as mean ± s.e.m. *$P < 0.05$ compared to control group. Additional details are given in Methods.

METHODS SUMMARY

For 4C-seq, chromatin was digested with DpnII and Csp6I. Captured DNA was amplified using promoter-specific primers and deep sequenced.

For 3C, nuclei were digested with HindIII. Primer quality was assessed using serial dilutions of BACs encompassing the regions of interest (RP23-268O10, RP23-96F3). The average of four independent experiments is represented graphically (Extended Data Fig. 2c).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 August 2013; accepted 10 February 2014.
Published online 12 March 2014.

1. Dina, C. *et al.* Variation in FTO contributes to childhood obesity and severe adult obesity. *Nature Genet.* **39,** 724–726 (2007).
2. Frayling, T. M. *et al.* A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316,** 889–894 (2007).
3. Scuteri, A. *et al.* Genome-wide association scan shows genetic variants in the *FTO* gene are associated with obesity-related traits. *PLoS Genet.* **3,** e115 (2007).
4. Church, C. *et al.* Overexpression of *Fto* leads to increased food intake and results in obesity. *Nature Genet.* **42,** 1086–1092 (2010).

5. Fischer, J. *et al.* Inactivation of the *Fto* gene protects from obesity. *Nature* **458,** 894–898 (2009).
6. Gao, X. *et al.* The fat mass and obesity associated gene *FTO* functions in the brain to regulate postnatal growth in mice. *PLoS ONE* **5,** e14005 (2010).
7. Grunnet, L. G. *et al.* Regulation and function of *FTO* mRNA expression in human skeletal muscle and subcutaneous adipose tissue. *Diabetes* **58,** 2402–2408 (2009).
8. Klöting, N. *et al.* Inverse relationship between obesity and *FTO* gene expression in visceral adipose tissue in humans. *Diabetologia* **51,** 641–647 (2008).
9. Wåhlén, K., Sjolin, E. & Hoffstedt, J. The common rs9939609 gene variant of the fat mass- and obesity-associated gene *FTO* is related to fat cell lipolysis. *J. Lipid Res.* **49,** 607–611 (2008).
10. McMurray, F. *et al.* Adult onset global loss of the *fto* gene alters body composition and metabolism in the mouse. *PLoS Genet.* **9,** e1003166 (2013).
11. Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactions in human cells. *Nature* **503,** 290–294 (2013).
12. Houweling, A. C. *et al.* Gene and cluster-specific expression of the *Iroquois* family members during mouse development. *Mech. Dev.* **107,** 169–174 (2001).
13. van Tuyl, M. *et al. Iroquois* genes influence proximo-distal morphogenesis during rat lung development. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **290,** L777–L789 (2006).
14. Gerken, T. *et al.* The obesity-associated *FTO* gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **318,** 1469–1472 (2007).
15. Qi, L. *et al.* Fat mass-and obesity-associated (*FTO*) gene variant is associated with obesity: longitudinal analyses in two cohort studies and functional test. *Diabetes* **57,** 3145–3151 (2008).
16. Stratigopoulos, G. *et al.* Regulation of *Fto/Ftm* gene expression in mice and humans. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **294,** R1185–R1196 (2008).
17. Ragvin, A. *et al.* Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to *HHEX*, *SOX4*, and *IRX3*. *Proc. Natl Acad. Sci. USA* **107,** 775–780 (2010).
18. Visel, A. *et al.* VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35** (Database issue), D88–D92 (2007).
19. Bosse, A. *et al.* Identification of the vertebrate Iroquois homeobox gene family with overlapping expression during early development of the nervous system. *Mech. Dev.* **69,** 169–181 (1997).
20. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
21. Gamazon, E. R. *et al.* Enrichment of *cis*-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. *Mol. Psychiatry* **18,** 340–346 (2012).
22. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genet.* **42,** 937–948 (2010).
23. Kong, D. *et al.* GABAergic RIP-Cre neurons in the arcuate nucleus selectively regulate energy expenditure. *Cell* **151,** 645–657 (2012).
24. Shi, Y. C. *et al.* Arcuate NPY controls sympathetic output and BAT function via a relay of tyrosine hydroxylase neurons in the PVN. *Cell Metab.* **17,** 236–248 (2013).
25. Mori, H. *et al.* Critical role for hypothalamic mTOR activity in energy balance. *Cell Metab.* **9,** 362–374 (2009).
26. Dankel, S. N. *et al.* Switch from stress response to homeobox transcription factors in adipose tissue after profound fat loss. *PLoS ONE* **5,** e11033 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank F. Gage, C. Marchetto, B. Ren and F. Jin for their generosity in sharing reagents and data. This work was funded by grants from the National Institutes of Health (DK093972, HL119967, HL114010 and DK020595) to M.A.N. and (MH101820, MH090937 and DK20595) to N.J.C. J.L.G.-S. was funded by grants from the Spanish Ministerio de Economía y Competitividad (BFU2010-14839, CSD2007-00008) and the Andalusian Government (CVI-3488). C.-C.H. was supported by a grant from the Canadian Institute of Health Research. K.-H.K. is supported by a fellowship from the Heart and Stroke Foundation of Canada. S.S. is supported by an NIH postdoctoral training grant (T32HL007381)

Author Contributions M.A.N., J.L.G.-S. and C.-C.H. designed the project. J.J.T. and C.G.-M. performed 4C-seq experiments, with analysis also aided by S.S. and N.J.S. N.J.S. performed the locus conservation and regulatory block analysis. I.A., F.L.C., D.R.S., N.F.W. and S.S. performed *in vivo* enhancer experiments. S.S. and M.S. performed *in situ* hybridizations and mouse knockout phenotype calculation. E.R.G. and N.J.C. performed eQTL analyses. K.-H.K. performed mouse metabolic experiments. J.H.L. and D.T. contributed to histological analysis and glucose homeostasis analysis. V.P. contributed to gene and protein expression analysis. J.E.S. contributed to metabolic cage analysis. H.K.S., D.R.S., M.M., S.N., N.A.V., R.D.A. and A.N. provided scientific discussion and technical support. S.S., C.C.H., J.L.G.-S. and M.A.N. wrote the paper with input from all authors.

Author Information Data were submitted to GEO under accession GSE52830. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.N. (nobrega@uchicago.edu), J.L.G.-S. (jlgomska@upo.es) or C.-C.H. (cchui@sickkids.ca).

METHODS

3C assays. 3C assays were performed as described in ref. 27. Mouse brain tissue was processed to get single cells samples. Ten-million isolated cells were treated with lysis buffer (10 mM Tris-HCl, pH 8, 10 mM NaCl, 0.3% IGEPAL CA-630 (Sigma-Aldrich, I8896), 1X protease inhibitor cocktail (cOmplete, Roche, 11697498001)) and nuclei were digested with HindIII endonuclease (Roche, 10798983001). Subsequently, DNA was ligated with T4 DNA ligase (Promega, M1804). A set of locus specific primers (see table) was designed with the online program primer3 v. 0.4.0 (ref. 28), each one close to a HindIII restriction site. These primers were used to make real-time quantitative PCRs using a BioRad CFX96 Real Time System with SsoFast EvaGreen Supermix (BioRad 172-5203), in order to measure the relative enrichment in each ligation product. The primer designed within the linkage disequilibrium-block was used as fixed primer. The quality of all primer pairs was measured using serial dilutions of a bacterial artificial chromosome (BAC) DNA that encompass the regions of interest (RP23-268O10, RP23-96F3). PCR values were normalized by means of control primers designed in the *Ercc3* gene locus. The average of four independent experiments is represented graphically (Extended Data Fig. 2c).

4C-seq. 4C-seq assays were performed as previously reported[27,29–31]. Whole embryos (9.5 days of development) or whole adult (8 weeks) mice brains (both strain CD-1; Charles River) were processed to get approximately 10-million isolated cells, which were treated with lysis buffer (10 mM Tris-HCl pH 8, 10 mM NaCl, 0.3% IGEPAL CA-630 (Sigma-Aldrich, I8896), 1X protease inhibitor cocktail (cOmplete, Roche, 11697498001)). Nuclei were digested with DpnII endonuclease (New England Biolabs, R0543M) and ligated with T4 DNA ligase (Promega, M1804). Subsequently, Csp6I endonuclease (Fermentas, Thermo Scientific, FD0214) was used in a second round of digestion, and the DNA was ligated again. Specific primers were designed near each gene promoter with primer3. Viewpoint fragend coordinates (for mouse genome build m9) were as follows: *Fto*, chromosome 8, 93,837,310–93,837,724; *Irx3*, chromosome 8, 94,325,277–94,326,009. Illumina adaptors were included in the primers sequence[32]. Sixteen PCRs were performed with Expand Long Template PCR System (Roche, 11759060001) for each viewpoint, pooled together and purified using High Pure PCR Product Purification Kit (Roche, 11732668001). Quanti-iT PicoGreen dsDNA Assay Kit (Invitrogen, P11496) was used in order to measure sample concentration, and then deep sequenced. 4C-seq data were analysed, with some changes, as described in ref. 30.

In brief, raw sequencing data were de-multiplexed and aligned using either the Mouse July 2007 (NCBI37/mm9) or Zebrafish December 2008 (Zv8/danRer6) assembly as the reference genome. Reads located in fragments flanked by two restriction sites of the same enzyme, or in fragments smaller than 40 bp were filtered out. Mapped reads were then converted to reads-per-1kb-bin units, and smoothed using a running mean 5-fragment window algorithm. To calculate the statistically significant targets for each viewpoint, a background theoretical model was calculated as the exponential fit curve of the average signal of 61 4C-seq different samples[11,33,34]. The P value for each bin was calculated by means of Poisson probability function. Smoothed data were uploaded to the UCSC Browser[35] for visualization and plotted using the browser and Circos[36]. For plotting, statistically significant targets were defined as $P < 0.01$.

Primers used for 3C experiments. The following primers were used for 3C experiments: '3c_mouse_LDBlock', 5′-TGGTCTCGGGTATCTTGTCC-3′; '3c_mouse_Irx3_control1', 5′-TCGCTCCTAGTGAGACTTTC-3′; '3c_mouse_Irx3_promoter', 5′-TGATGTTGGTTCCTTACTAGG-3′; '3c_mouse_Irx3_control2', 5′-TCCACAGAAACATCTGACG-3′; '3c_mouse_Ercc3-1', 5′-TGACCTCCACACTCCTGAC-3′; '3c_mouse_Ercc3-2', 5′-ATGCGCAATTAGAAACTGC-3′.

Primers used for 4C-seq experiments. The following primers were used for 4C-seq experiments: '4c_mouse_Fto_read', 5′-AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCC GATCTCATATTGCTCTGGATGCAGATC-3′; '4c_mouse_Fto_noread', 5′-CAAGCAGAAGACGGCATACGAGAACTACCCTTCCCAGAATGC-3′; '4c_mouse_Irx3_read', 5′-AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGAT CTCCGCCGCGGAGCAGATC-3′; '4c_mouse_Irx3_noread', 5′-CAAGCAGAAGACGGCATACGATGGAGTCGCCAATCACC-3′; '4c_zebrafish_Fto_read', 5′-AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGAT CTCCTCCACTGTCATCCGATC-3′; '4c_zebrafish_Fto_noread', 5′-CAAGCAGAAGACGGCATACGACTACATTTTATGTCAGTTTCGGG-3′; '4c_zebrafish_Irx3a_read', 5′-AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGAT CTCCTACCGGATTACTCTACAGATC-3′; '4c_zebrafish_Irx3a_noread', 5′-CAAGCAGAAGACGGCATACGAAAAACGCCCAGAAGACTGAC-3′.

ENCODE ChIA-PET. Publicly available data for the breast cancer MCF7 cell line, generated as part of the ENCODE project[37], were downloaded from and visualized with the WashU EpiGenome Browser[38] (http://epigenomegateway.wustl.edu/browser/). ChIA-PET has been performed to capture long-range chromatin interactions associated with RNA polymerase II[39].

Locus conservation. To assess the level of high-level conservation of the *FTO-IRX5* locus, we used phastCons[40] elements and scores as a measurement of conservation.

High-scoring phastCons elements (LOD > 680, top 1%) cover 3% (38,040 out of 1,244,409 bp) of the genomic space that includes *FTO* and *IRX5* (hg19, chromsome 16: 53,731,249–54,975,288), whereas on average, only 0.5 ± 0.6% of genomic intervals with the same length are occupied by such high-scoring phastCons elements. Genomic intervals of the same size of the *FTO-IRX3* locus were generated every 10 kb for each chromosome. The *FTO-IRX5* is at the top 1% intervals containing high-scoring phastCons elements.

Similarly, in the *FTO-IRX5* interval, 9% of the base pairs have phastCons score above 0.9 (range: 0.0–1.0), whereas 14,164 intervals from all chromosomes of the same size offset by 200 kb contain only 3.8% ± 1.9 of such high-scoring base pairs. At 9%, the *FTO-IRX5* is at the top 2% intervals with the most conserved base pairs.

These results indicate that the *FTO-IRX5* interval contains significantly more highly conserved base pairs than most genomic loci.

Mouse *in vivo* transgenic reporter assays. Enhancer reporter assays and constructs, including the modified BAC, were created as reported previously[41]. The human (NCBI37/hg19) genomic coordinates of the elements cloned are: FTO-1, chromosome 16, 53,800,270–53,805,184; FTO-2, chromosome 16, 53,808,432–53,813,121; FTO-3, chromosome 16, 53,816,312–53,821,148. FTO-promoter, chromosome 16, 53,736,860–53,738,096; IRX3-promoter, chromosome 16, 54,319,963–54,322,782. BAC RP11-261B9 was modified by inserting a *LacZ-Amp* cassette in place of the 36 nucleotides in exon 1 at chromosome 16, 53,738,097–53,738,132. This homologous recombination was facilitated by using the following homology arms: FTO-hArm1, chromosome 16, 53,738,047–53,738,096; FTO-hArm2, chromosome 16, 53,738,133–53,738,182.

Mouse *in vivo* transgenic reporter assays were performed as reported previously[41]. *In situ* hybridization assays were performed according to standard protocols[42]. The full-length mouse Fto coding DNA sequence (CDS) as riboprobe template was from IMAGE clone 5708558 (ATCC).

Gene expression in human tissues. For the eQTL analyses, we used cerebellum[21] ($n = 153$; publicly available in GEO under accession number GSE35974) and adipose[43] ($n = 62$; publicly available in GEO under accession number GSE40234); details of the samples, the quality control procedure and the normalization approach used were described previously[21]. For adipose, the samples were chosen to be at the tails of the distribution of insulin sensitivity matched on age, gender and natural log of BMI. Because the adipose samples included individuals of African ancestry, we performed principal component analysis and tested each marker in the locus for the additive effect of allele dosage on the residual expression phenotype (IRX3 and FTO) after adjusting for the principal components ($n = 2$). Similarly, for cerebellum, we performed linear regression on the residual expression phenotype (IRX3 and FTO) after adjusting for sex and pH to evaluate the additive effect of allele dosage at each marker in the locus; the non-diseased samples (obtained from the Stanley Medical Research Institute) included in the analysis were of European descent and thus the principal components were not used to generate the residual expression. Significance (P value) was evaluated using the t-statistic from the regression.

Single nucleotide polymorphisms associated with gene expression levels. We downloaded genome-wide association data from the Genetic Investigation of Anthropometric Traits[22] (GIANT) Consortium website (http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files#BMI_.28download_GZIP.29). We extracted the P values for those SNPs in the locus with significant association ($P < 0.05$) with IRX3 or FTO expression in adipose or cerebellum. We plotted the distribution of association P values for these expression-associated SNPs (eSNPs). rs9930506 is highly associated with BMI ($P = 1.41 \times 10^{-53}$, $n = 123541$). Recombination rates are estimated from the International HapMap Project.

Statistical analyses. Statistical analyses were carried out using the R statistical software (http://cran.r-project.org/). Regional plots were done by repurposing an R script retrieved from the Broad Institute (http://www.broadinstitute.org/diabetes/scandinavs/figures.html).

Mice. All animal experimental protocols approved by the Animal Care Committee of the Toronto Centre of Phenogenomics conformed to the standards of the Canadian Council on Animal Care. *Irx3*-deficient (*Irx3*-knockout) mice[44] and *Ins2-Cre* mice[23] were described previously. $Rosa26^{EnR\text{-}Irx3}$ conditional transgenic mice (*EnR-Irx3*) were generated using pROSA26PA gene-targeting vector as described previously[45]. Mice were maintained on 12-h light–dark cycles and provided with food and water *ad libitum*. For diet-induced obesity studies, 8-week-old male mice were subjected to 45% high-fat diet (Research Diets) for 10 weeks. Body weight was measured every week from 4 to 18 weeks of age, and body composition was analysed using an EchoMRI device (Echo Medical Systems) at 18 weeks of age. Body mass index (BMI) was calculated by dividing body weight (g) by body length (mm) squared (BMI = body weight/body length^2).

Metabolic phenotyping experiments. Energy expenditure was evaluated through indirect calorimetry (Oxymax System, Columbus Instruments) over periods of 24 h.

In brief, energy expenditure was calculated by multiplying oxygen consumption (VO_2) by the calorific value (CV = 3.815 + 1.232 × respiratory exchange ratio) and corrected by lean mass. Locomotor activity and food intake were also measured simultaneously. For glucose and insulin tolerance tests, mice were subjected to intraperitoneal bolus injection of glucose (1 mg g^{-1} of body weight) or insulin (1.5 milli-units per g) after fasting for overnight (14 to 16 h) or 6 h with water *ad libitum*, respectively. Blood glucose levels were measured at the indicated intervals. For histological analysis, tissues were fixed in 4% paraformaldehyde and embedded in paraffin. Sections of 5 μm were stained with haematoxylin and eosin (H&E). *Irx3* expression in brain and lung was examined using *Irx3^{lacZ}* knock-in mice with β-galactosidase staining.

Gene and protein expression analyses. Total RNA was extracted from fat tissues or hypothalamus using RNeasy Lipid Tissue Kit (Qiagen), and complementary DNA was synthesized from 2 μg of RNA using M-MLV reverse transcriptase (Invitrogen) with oligo(dT). Gene expression assay was conducted using SYBR Green methods on Viia7 (Applied Biosystems), and relative cycle threshold (CT) values were normalized by β-actin. Primer sequences are available upon request. Western blot analysis was performed as described previously[46]. The same amount of protein extracted from hypothalamus or brown adipose tissues was loaded, which was confirmed by β-actin as a loading control. Antibodies used are as follows: Ucp1 (Santa Cruz Biotechnology), Irx3 (generated in-house) and β-actin (Calbiochem).

Statistics. All results are expressed as mean ± s.e.m. Statistical significance of differences among groups was determined by Student's *t*-test or analysis of variance (ANOVA) with post-hoc analysis, Student–Newman–Keuls using Sigma Stat (SPSS). Differences at $P < 0.05$ were considered statistically significant (*$P < 0.05$ versus WT or control; †$P < 0.05$ versus WT, HFD). The following numbers of mice per sample were used for comparison; body-weight measurement for *Irx3* mice (ND, WT/knockout (KO), $n = 32/25$; HFD, WT/KO, $n = 12/7$) and *EnR-Irx3* mice (control/mutant, $n = 9/12$); body composition analysis for *Irx3* mice (ND WT/KO, $n = 21/12$; HFD WT/KO, $n = 8/5$) and *EnR-Irx3* mice (control/mutant, $n = 9/12$); glucose tolerance test for *Irx3* mice (ND 8 weeks, WT/KO, $n = 8/5$; ND 18 weeks, WT/KO, $n = 10/6$; HFD 18 weeks, WT/KO, $n = 12/7$) and *EnR-Irx3* mice (control/mutant, $n = 5/8$); insulin tolerance test for *Irx3* mice (WT/KO, $n = 12/7$); indirect calorimeter analysis for *Irx3* mice(ND WT/KO, $n = 7/5$; HFD WT/KO, $n = 8/4$) and *EnR-Irx3* mice (control/mutant, $n = 5/7$), and gene expression analysis for *Irx3* mice (WT/KO, $n = 10/7$) and *EnR-Irx3* mice (control/mutant, $n = 5/7$).

Mouse body-mass phenotypes. To calculate the fraction of knockout mice that display a phenotype affecting body mass, we consulted the Mouse Genome Informatics (MGI) database[47].

First, to determine the number of gene knockouts that result in a body mass phenotype, we needed to define a more precise phenotype. The Mammalian Phenotype Ontology[48] (http://www.informatics.jax.org/searches/MP_form.shtml) specifies a controlled vocabulary by which all mice in the MGI database are classified. Within this framework, (we believe) the most applicable is 'abnormal postnatal growth/weight/body size' (MP:0002089).

Next, we crafted a query using the MouseMine portal[49] (http://www.mousemine.org) to access the MGI database, using MP:0002089 as the search criteria in the 'Mammalian phenotypes → Mouse genotypes' template (http://www.mousemine.org/mousemine/template.do?name=MPhenotype_MFeature&scope=all). We asked for all targeted alleles on chromosomes 1–19, X and Y, reasoning that many non-targeted (that is, random, chemical and radiation-induced) mutations have not been precisely mapped, making it difficult if not impossible to ensure that they are not counted multiple times. Also, in the context of our study, we are discussing the chance that a targeted knockout would give a specific phenotype, so these types of alleles are most relevant. Finally, we excluded alleles present only as cell lines, which left 2,166 unique genes.

To determine the total number of targeted mutations present in the MGI database, we performed a search identical to that above but omitting the MP:0002089 term, which left 7,556 unique genes.

Genomic regulatory blocks. Genomic regulatory blocks are regions under strong selection that contain a number of syntenic highly conserved sequences, believed to regulate a syntenic target gene ('anchor gene', the gene that is the raison d'être of the GRB). There are different approaches to determine GRBs, usually identifying syntenic genes and non-coding conserved regions that span a region and determining the boundaries of the region.

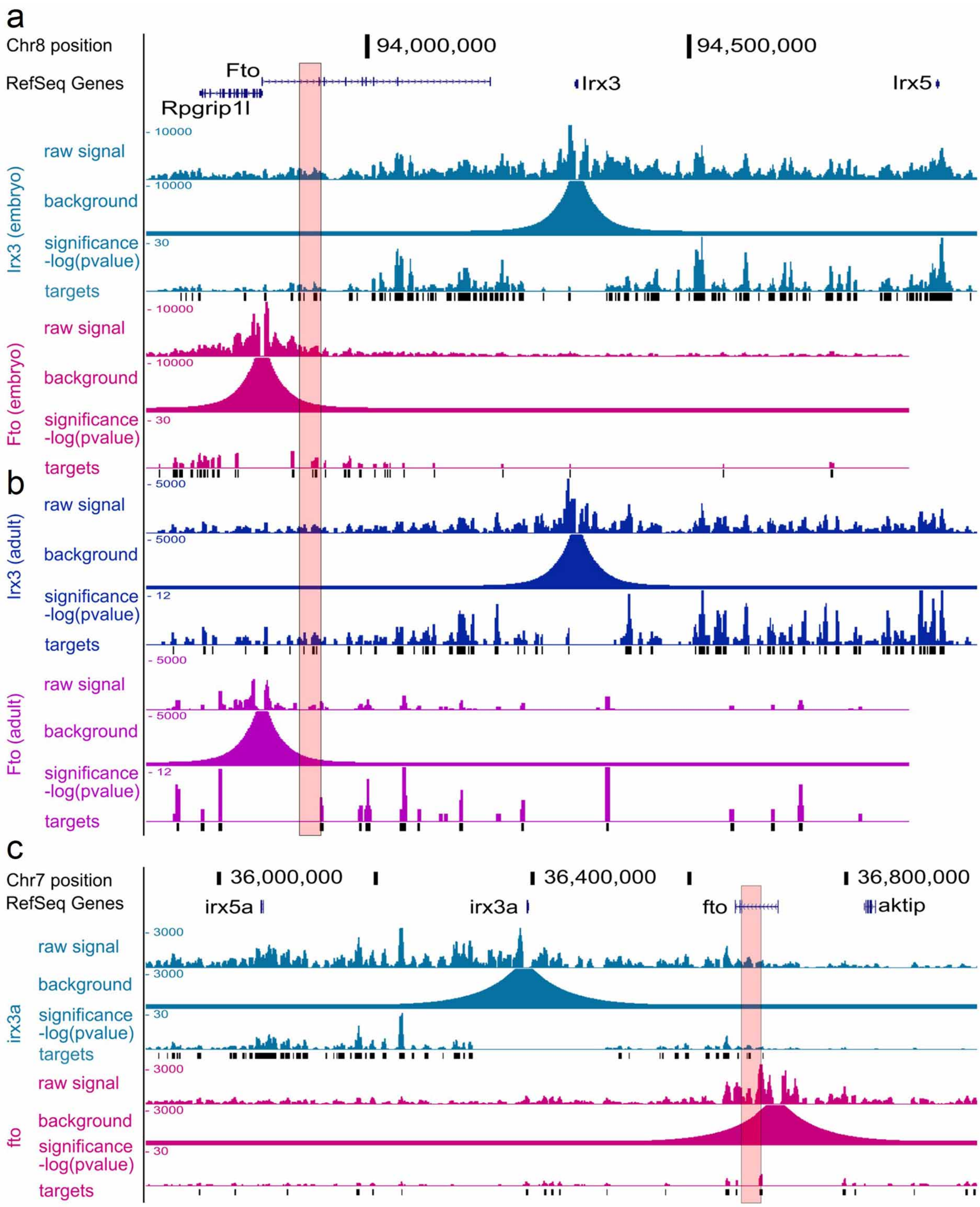
GWAS identified SNPs that might affect noncoding regulatory elements that, in turn, regulate specific target genes. Assignment of regulatory elements to targets is still a challenge, the simplest method being to assign them to the nearest gene. In the case of FTO, because the GWAS SNP occurs in an FTO intron, this gene was considered to be the target.

The idea put forward here is that when regulatory elements occur within a GRB, which is the case for FTO, IRX3 and IRX5, it is likely that they target the anchor gene (*IRX3* or *IRX5*), given its importance. Supplementary Table 5 contains an annotation of GRBs and their anchor genes, and the GWAS SNPs within the GRB, as well as annotation of whether the nearest gene is the anchor gene or not. In those cases in which the nearest gene differs from the anchor gene, we propose that the anchor gene be used as the putative target and additional experimentation as shown in this work should be performed to determine precisely which is the target gene for the GWAS-affected regulatory element.

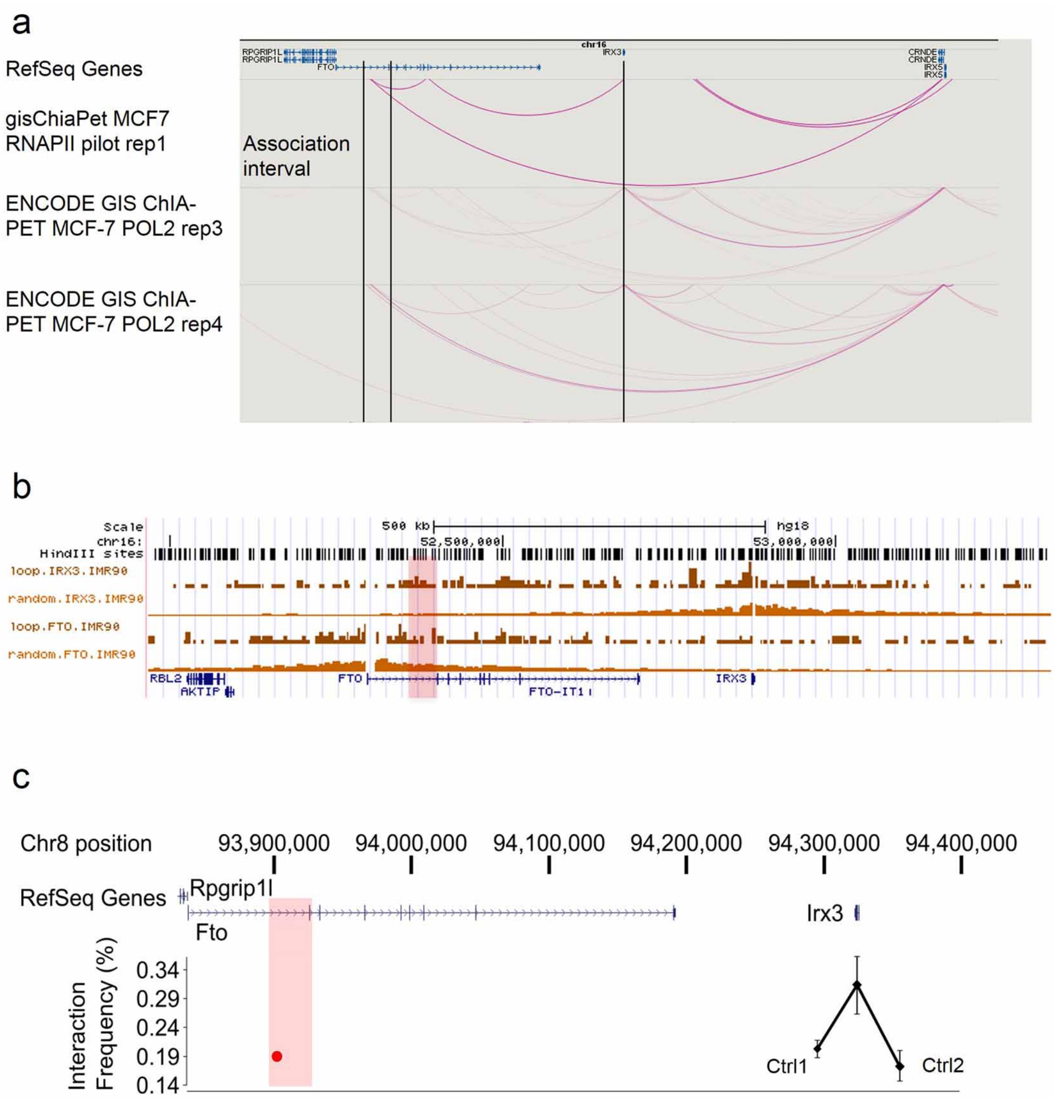
Genomic regulatory blocks were obtained from UCNEbase (http://ccg.vital-it.ch/UCNEbase/)[50]. Human UCNE clusters (hg19) were downloaded from http://ccg.vital-it.ch/UCNEbase/data/download/clusters/hg19_clusters_coord.bed and are referred here as GRBs.

The hg19 GWAS catalogue from the UCSC Genome Browser was downloaded and the tagging SNPs contained in GRBs with more than one gene (making SNP assigning doubtful) were kept for further annotation. SNPs occurring in exons were removed. The nearest gene was considered to be the RefSeq gene whose transcription start site was the nearest, or the gene where the SNP is contained.

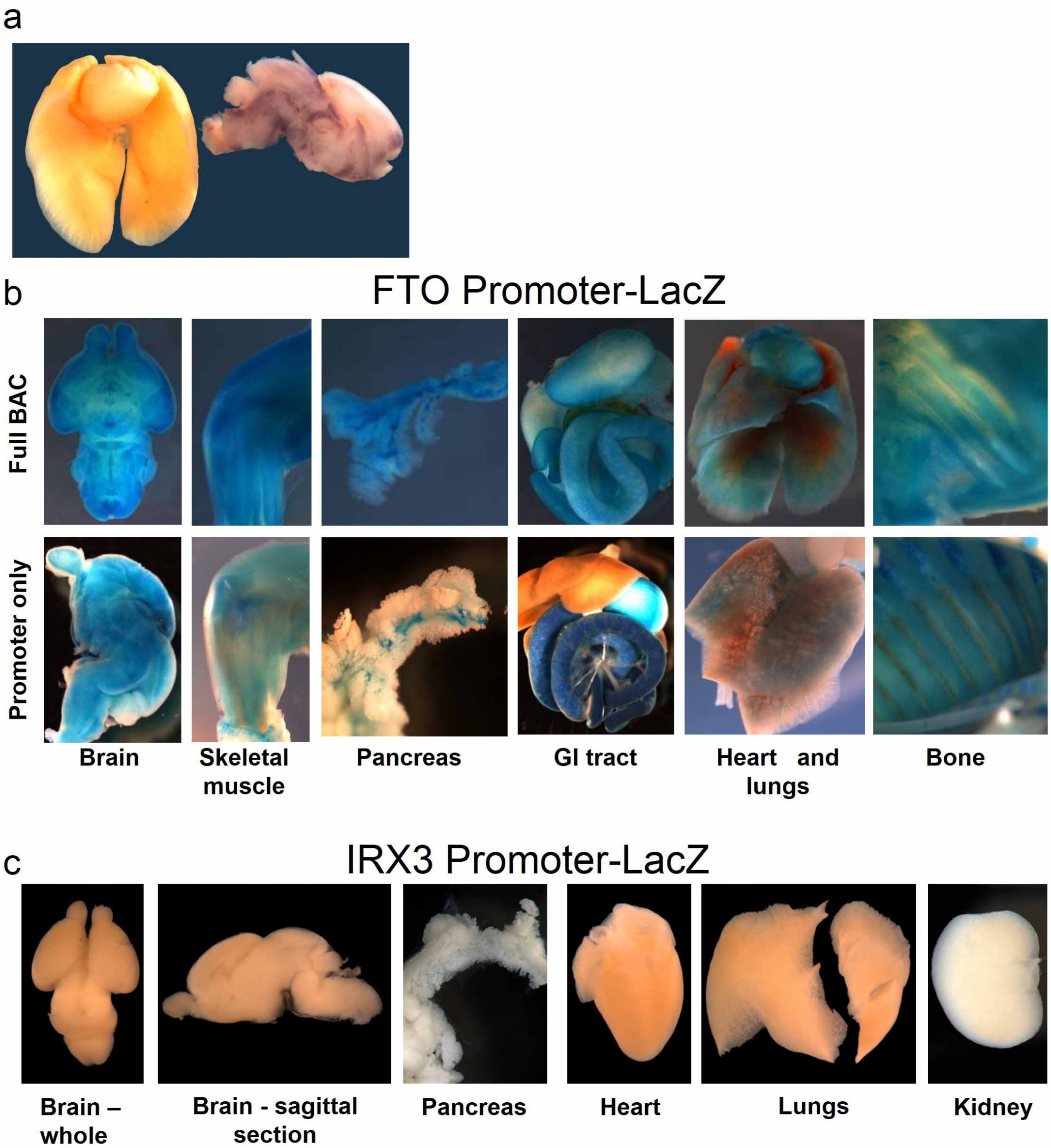
27. Hagège, H. *et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nature Protocols* **2,** 1722–1733 (2007).
28. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Meth. Mol. Biol.* **132,** 365–386 (2000).
29. Dekker, J. *et al.* Capturing chromosome conformation. *Science* **295,** 1306–1311 (2002).
30. Noordermeer, D. *et al.* The dynamic architecture of *Hox* gene clusters. *Science* **334,** 222–225 (2011).
31. Splinter, E. *et al.* Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* **58,** 221–230 (2012).
32. Stadhouders, R. *et al.* Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nature Protocols* **8,** 509–524 (2013).
33. Denholtz, M. *et al.* Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell* **13,** 602–616 (2013).
34. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).
35. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12,** 996–1006 (2002).
36. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19,** 1639–1645 (2009).
37. ENCODE. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9,** e1001046 (2011).
38. Zhou, X. & Wang, T. Using the Wash U Epigenome Browser to examine genome-wide sequencing data. *Curr. Protoc. Bioinformatics* **40,** 10.10.1–10.10.14 (2012).
39. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148,** 84–98 (2012).
40. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15,** 1034–50 (2005).
41. Smemo, S. *et al.* Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum. Mol. Genet.* **21,** 3255–3263 (2012).
42. Wilkinson, D. G. & Nieto, M. A. Detection of messenger RNA by *in situ* hybridization to tissue sections and whole mounts. *Methods Enzymol.* **225,** 361–373 (1993).
43. Elbein, S. C. *et al.* Genetic risk factors for type 2 diabetes: a *trans*-regulatory genetic architecture? *Am. J. Hum. Genet.* **91,** 466–477 (2012).
44. Zhang, S. S. *et al. Iroquois homeobox gene 3* establishes fast conduction in the cardiac His-Purkinje network. *Proc. Natl Acad. Sci. USA* **108,** 13576–13581 (2011).
45. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of *EYFP* and *ECFP* into the *ROSA26* locus. *BMC Dev. Biol.* **1,** 4 (2001).
46. Li, Z. J. *et al.* Kif7 regulates Gli2 through Sufu-dependent and -independent functions during skin development and tumorigenesis. *Development* **139,** 4152–4161 (2012).
47. Eppig, J. T. *et al.* The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. *Nucleic Acids Res.* **40** (Database issue), D881–D886 (2012).
48. Smith, C. L. & Eppig, J. T. The mammalian phenotype ontology: enabling robust annotation and comparative analysis. *Biol. Med.* **1(3)**, 390–399 (2009).
49. Smith, R. N. *et al.* InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data. *Bioinformatics* **28,** 3163–3165 (2012).
50. Dimitrieva, S. & Bucher, P. UCNEbase–a database of ultraconserved non-coding elements and genomic regulatory blocks *Nucleic Acids Res.* **41** (Database issue), D101–D109 (2013).



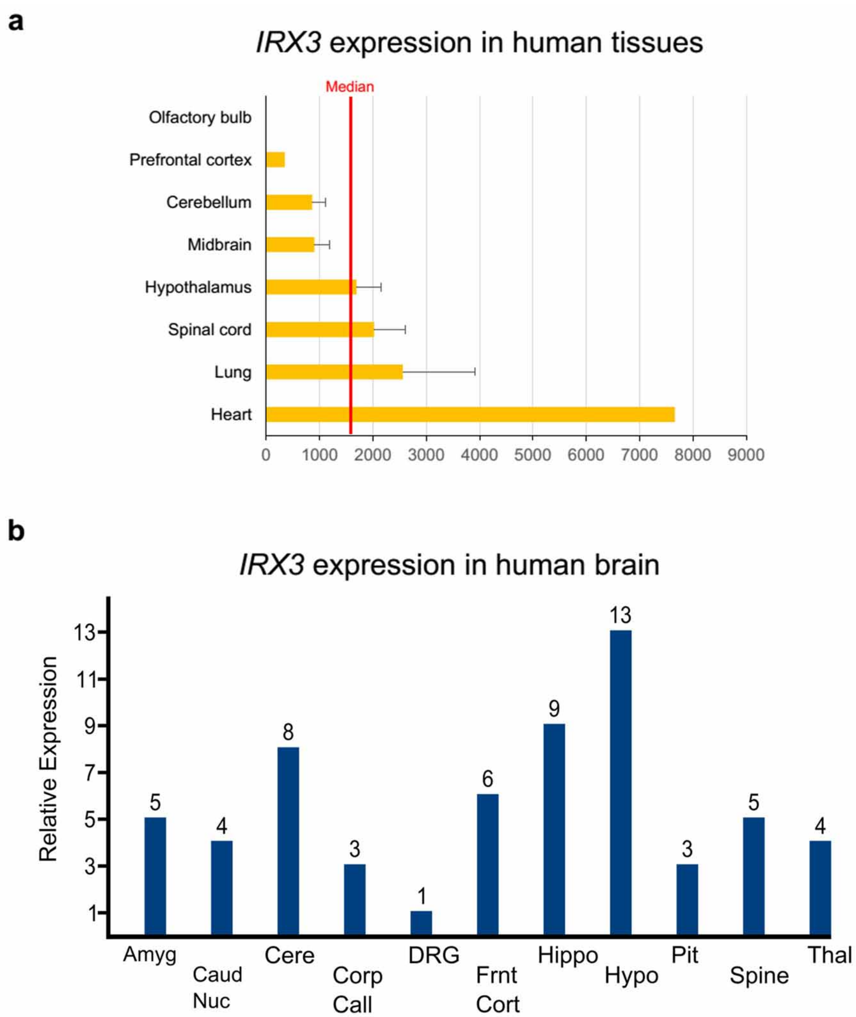
Extended Data Figure 1 | Long-range interactions in mouse and zebrafish. 4C-seq data for the *Fto*-*Irx3* locus, visualized with the UCSC Genome Browser. **a**, Data (also shown in the circular plot in Fig. 1) generated using whole mouse embryos (E9.5), showing the frequency of interactions with the promoter of *Irx3* (blue, top) or *Fto* (magenta, bottom). The background signal corrects for the strong correlation between (nonspecific) ligation events and the linear distance along the chromosome. Poisson statistical significance ($-\log(P$ value)) of the 4C-seq interactions over the background is plotted. Significant interactions ($P < 0.01$), 'targets', are displayed in black. **b**, As above for **a** but for adult mouse brain (8 weeks). **c**, As above for **a** but for whole zebrafish embryos (24 h post fertilization). In all, the region orthologous to the obesity association interval in the first intron of *Fto* is highlighted in pink.



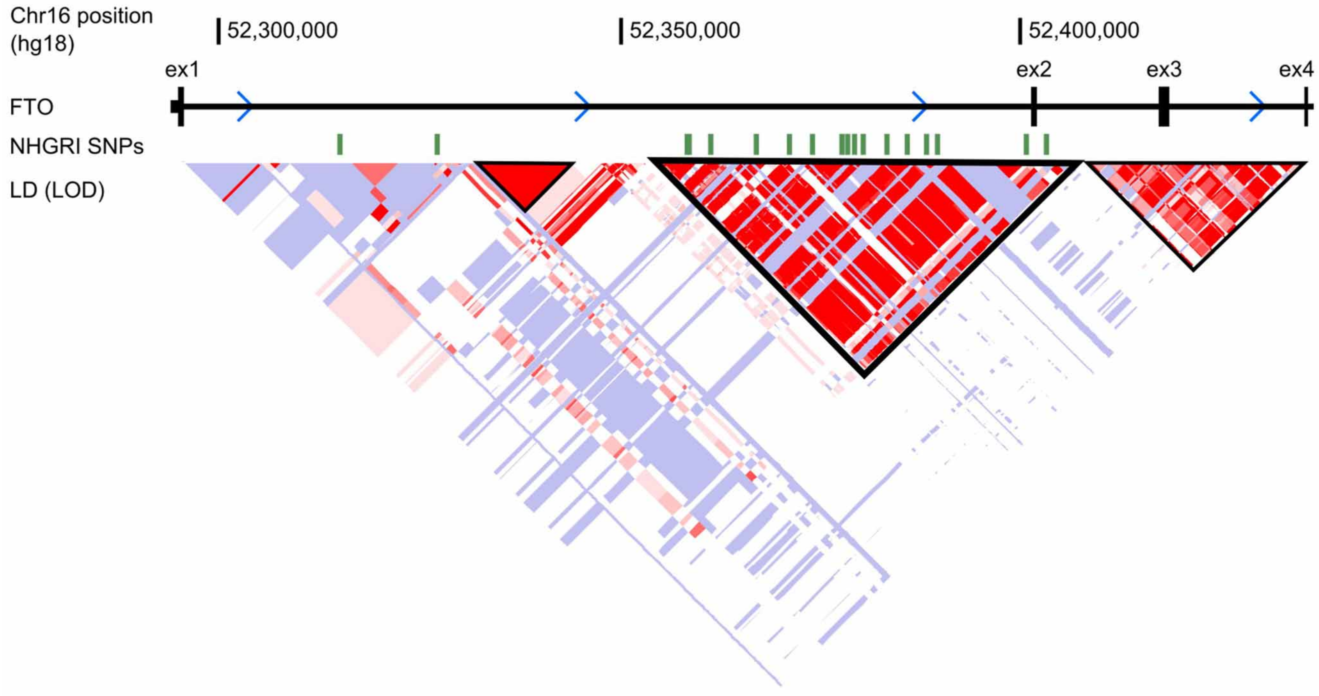
Extended Data Figure 2 | Long-range interactions at the *FTO-IRX3* locus. **a**, ENCODE data for ChIA-PET using RNA polymerase 2 (POL2) in MCF7 (human breast adenocarcinoma) cells shows interactions between *IRX3* and the obesity association interval in the first intron of *FTO*. No interactions are observed between the *FTO* promoter and the association interval. These public data are available from and was visualized with the WashU EpiGenome Browser (http://epigenomegateway.wustl.edu/browser/). **b**, Hi-C data previously generated[11] in human IMR-90 (fetal lung) cells. In the association interval, the *IRX3* signal is stronger than the background (random) signal. However, the signal for *FTO* is not. **c**, 3C data generated with adult (8 weeks) mouse brain. Using bait (red circle) in the association interval (red rectangle), we observe more frequent interactions with the *Irx3* promoter compared to control regions 1 and 2 that are 29 and 42 kb away, respectively, indicative of looping.



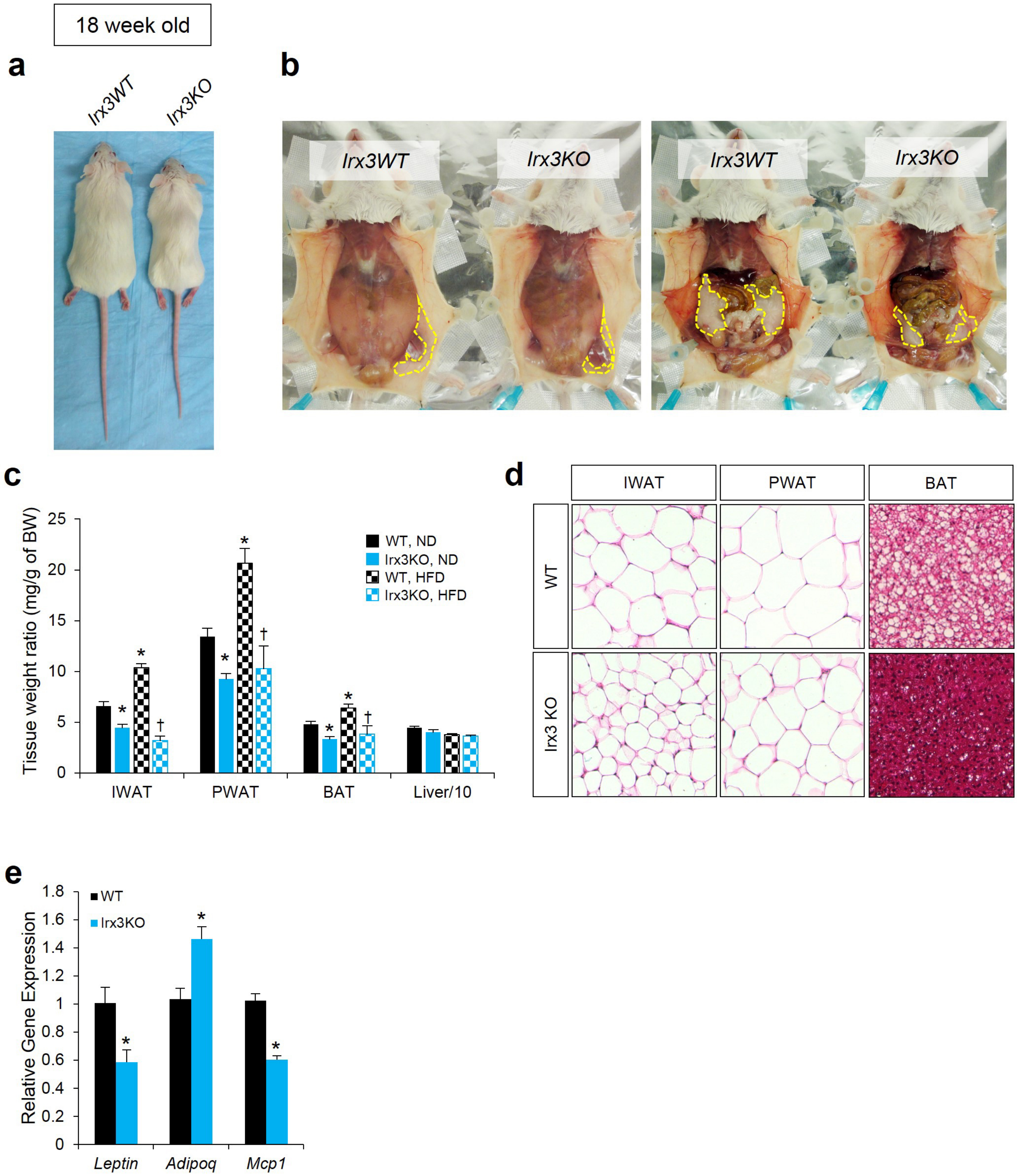
Extended Data Figure 3 | Gene expression in mouse tissue. **a**, *FTO* expression in lung and brain, shown by RNA *in situ* hybridization for mouse *Fto* mRNA, in newborn (P1) mouse. Lungs and heart (left, whole organs) were processed simultaneously and in the same well as brain (right, sagittal section) so that the relatively higher expression in brain can be observed. **b**, LacZ staining for β-galactosidase expression driven from the human *FTO* promoter. Top, the promoter–LacZ fusion is in the context of 162-kb of human genomic sequence carried in a BAC containing the first three exons of *FTO*, the entire obesity-associated interval and any enhancers present. The broad expression is consistent with previous reports in human and mouse (see main text for references). At bottom, the promoter–LacZ construct is isolated: only the 1,237 bp proximal to the transcriptional start site are included. Broad expression is recapitulated, indicating the robust transcriptional competency of the human *FTO* promoter. **c**, In contrast, the 2,820-bp proximal human *IRX3* promoter is not sufficient to drive LacZ expression, which is consistent with an enhancer-dependent transcriptional control mechanism.



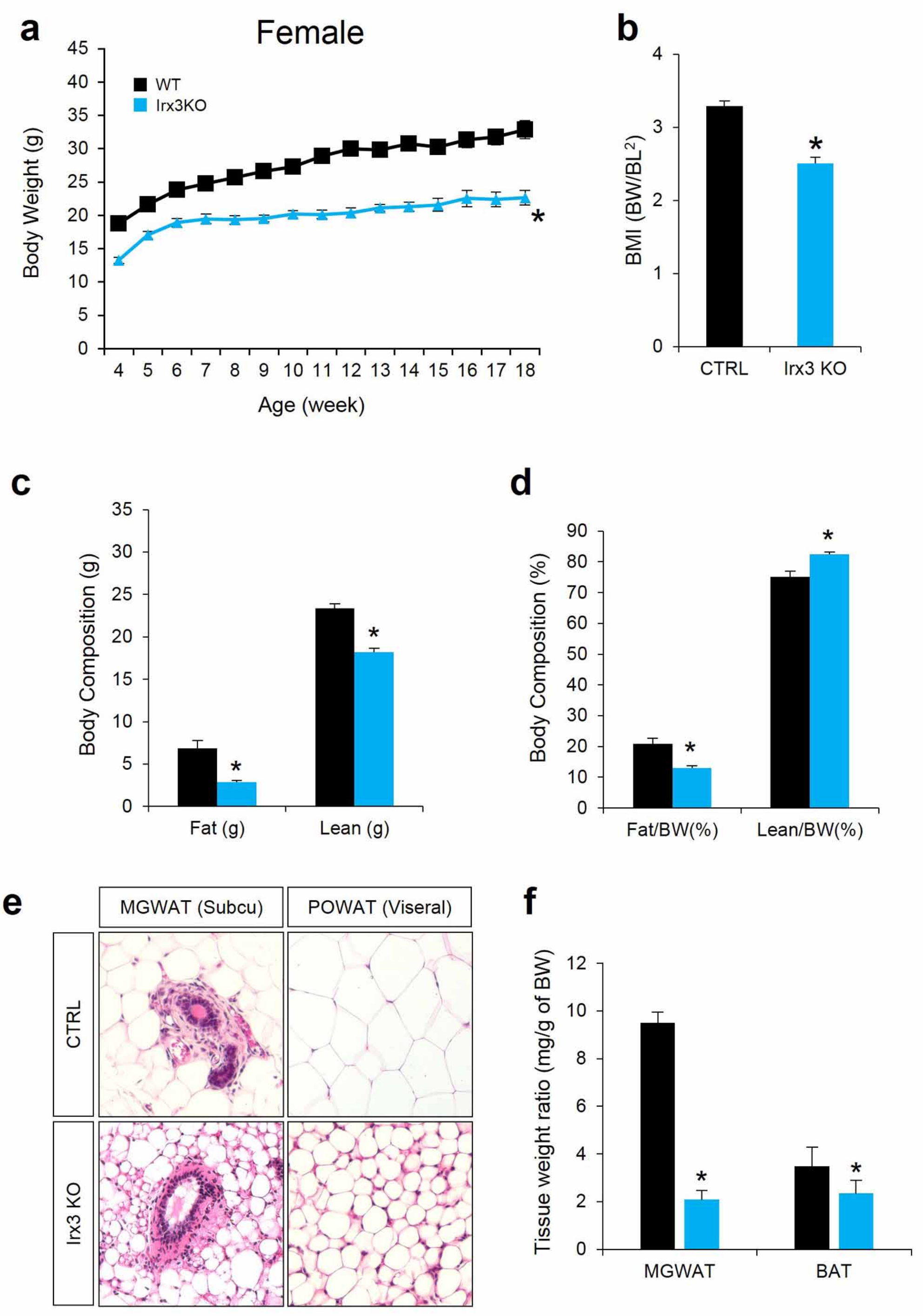
Extended Data Figure 4 | *IRX3* expression in human brain. **a**, *IRX3* expression in human tissues including brain. Expression data, measured on Affymetrix HG-U133 arrays, were obtained from the Body Atlas, Tissues (http://www.nextbio.com). The median expression across all 128 human tissues from 1,068 arrays is shown by the red line. **b**, *IRX3* expression in 11 different regions of human brain. Data were retrieved from Human Brain Transcriptome data (http://www.molecularbrain.org). Amyg: amygdala; Caud nuc: caudate nucleus; Cere: cerebellum; Corp Call: corpus callosum; DRG: dorsal root ganglion; Frnt Cort: frontal cortex; Hippo: hippocampus; Hypo: hypothalamus; Pit: pituitary; Spine: spinal cord; Thal: thalamus.



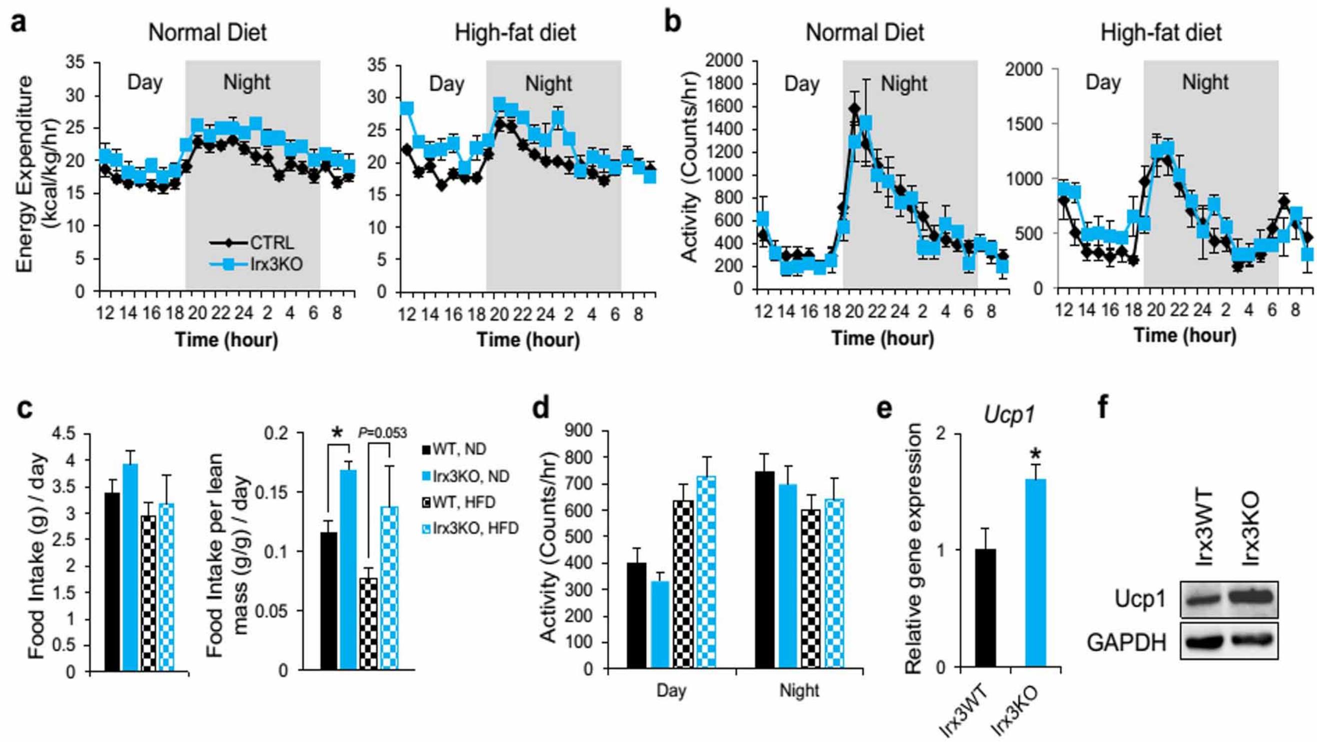
Extended Data Figure 5 | Linkage disequilibrium in *FTO*. Linkage disequilibrium (LD) plot of (logarithm (base 10) of odds) (LOD) score from HapMap phase II European data set, visualized in the UCSC browser. LD blocks are outlined in black. Obesity-associated SNPs from the National Human Genome Research Institute (NHGRI) GWAS catalogue are shown above, in green, demonstrating why this LD block is considered to define the 'association interval'.



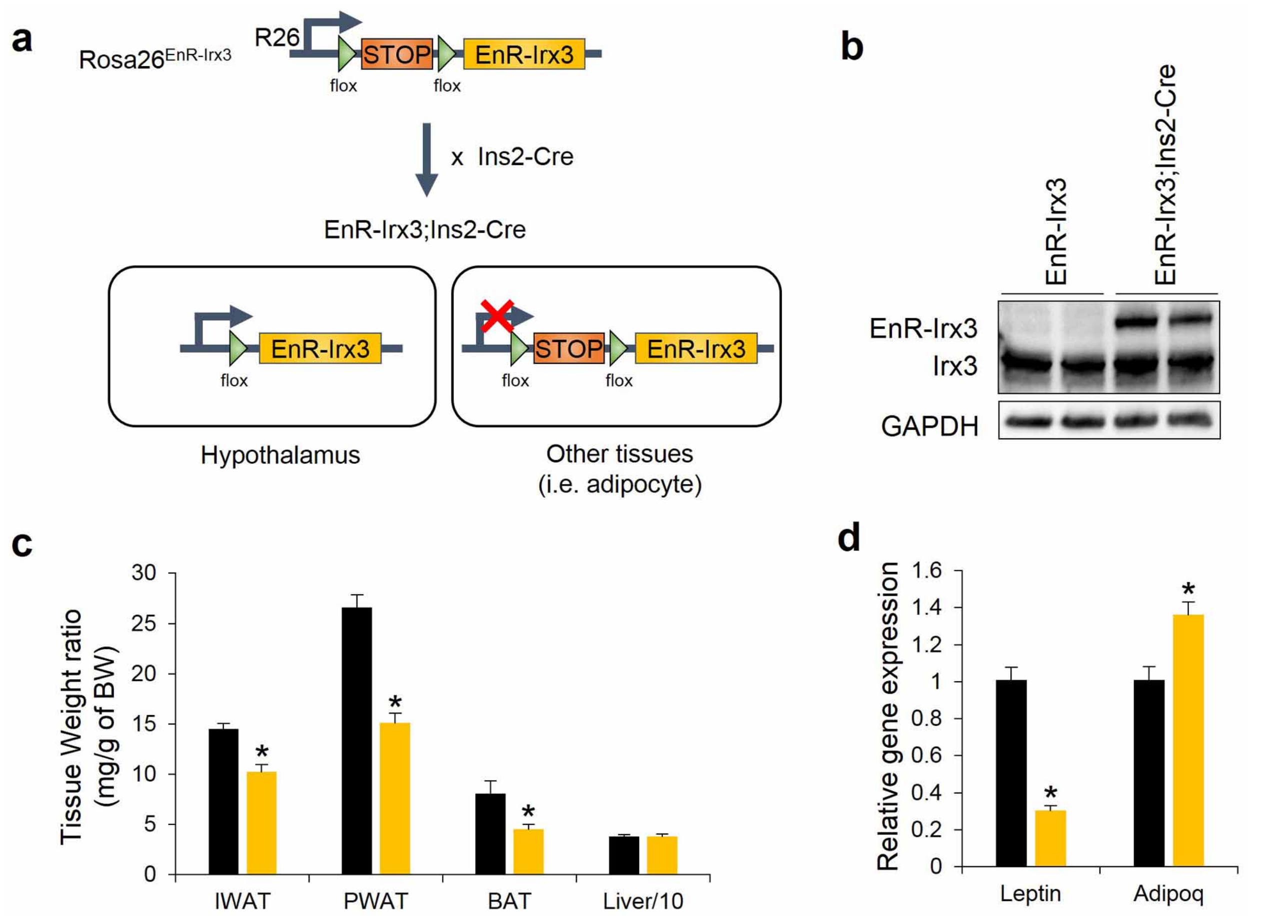
Extended Data Figure 6 | *Irx3*-knockout male mice are leaner with reduced adiposity. **a**, Representative photograph of WT and *Irx3* KO mice fed ND at 18 weeks of age. **b**, Representative anatomical views of WT and *Irx3* KO mice fed ND. Yellow dotted lines depict subcutaneous IWAT (left) and visceral PWAT (right). **c**, Tissue weights as a percentage of body weight showed smaller fat pad sizes in *Irx3* KO mice, compared to WT mice, in both ND and HFD conditions. (ND, WT/KO, $n = 20/12$; HFD, WT/KO, $n = 8/5$.) Data are mean ± s.e.m. (*$P < 0.05$ versus WT, ND; †$P < 0.05$ versus WT, HFD). **d**, Representative H&E sections of PWAT, IWAT and BAT from ND mice demonstrated smaller adipocyte size in *Irx3* KO mice than control mice. **e**, Quantitative PCR of WT and *Irx3* KO PWAT for the indicated marker genes: leptin (*lep*) and adiponectin (*adipoq*) are adipogenic markers, positively and negatively associated with adiposity, respectively; *Mcp1* correlates positively with adiposity. (*$P < 0.05$ versus WT value.) (WT/KO, $n = 10/7$).



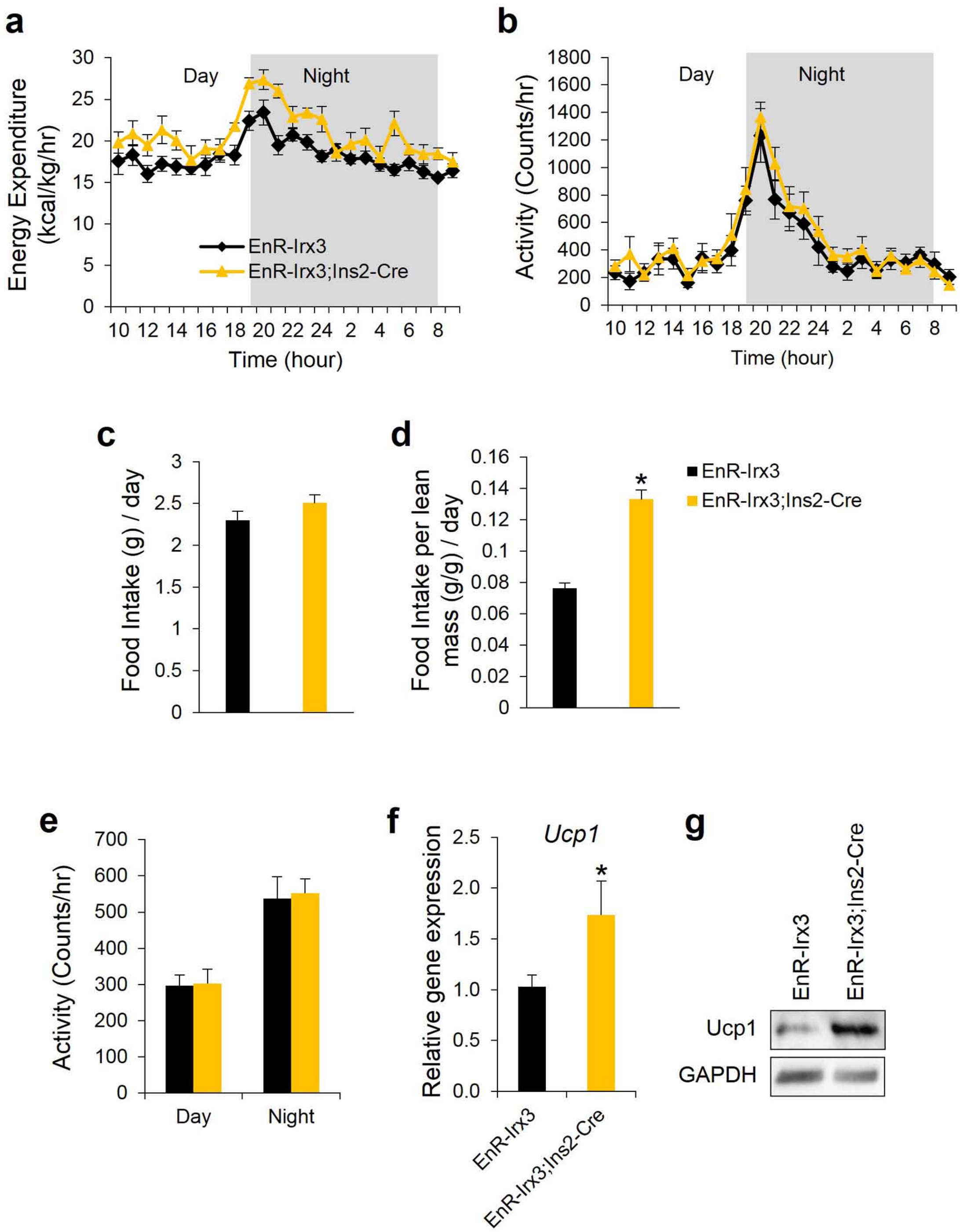
Extended Data Figure 7 | *Irx3*-knockout female mice are leaner with reduced adiposity. **a**, Body weight (BW) changes of WT and *Irx3* KO female mice fed a normal diet (ND). (WT/KO: $n = 15/14$). **b**, BMI, calculated by BW/BL^2 (BL, body length), is lower in *Irx3* KO female mice. (WT/KO, $n = 7/7$). **c**, **d**, Body composition analysis showed reduced fat mass and to a lesser extent reduced lean mass in *Irx3* KO female mice compared to WT mice, leading to decreased fat mass ratio (WT/KO, $n = 9/8$). **e**, Representative H&E-stained sections of mammary gland (MG) WAT and periovarian (PO) WAT revealed smaller adipocyte size in *Irx3* KO female mice, compared to WT. **f**, MGWAT and BAT weights as a percentage of body weight (WT/KO, $n = 4/5$). Data are mean ± s.e.m. (*$P < 0.05$ versus WT value.)



Extended Data Figure 8 | Higher energy expenditure of *Irx3*-knockout mice. **a**, Energy expenditure over a 24-h period, corrected for lean mass (kcal kg^{-1} h^{-1}), for 18-week-old WT and *Irx3* KO mice fed with ND and HFD (ND WT/KO, *n* = 7/5; HFD WT/KO, *n* = 8/4). **b**, Locomotor activity of WT and *Irx3* KO mice. **c**, Average amount of food intake over a 24-h period with or without normalization to lean mass. **d**, Average locomotor activity measured over 24 h. **e, f**, Elevated Ucp1 gene and protein expression in BAT (WT/KO, *n* = 7/6). Data are mean ± s.e.m. $*P < 0.05$ versus WT value.



Extended Data Figure 9 | Hypothalamic-specific *Irx3* dominant-negative mice are leaner with reduced adiposity. a, Schematic diagram of generation of transgenic mice overexpressing dominant-negative *Irx3* in the hypothalamus. **b**, Immunoblot analysis showed *EnR-Irx3* expression in the hypothalamus of mutant mice without affecting endogenous Irx3 expression, compared to control mice. **c**, Tissue weights as a percentage of body weight showed that fat pad sizes are smaller in mutant mice, compared to control mice. **d**, Reduced leptin expression and increased adiponectin gene expression in PWAT of mutant mice (control/mutant, $n = 5/7$). Data are expressed as mean ± s.e.m. $*P < 0.05$ compared to control group.



Extended Data Figure 10 | Higher energy expenditure of hypothalamic dominant-negative *Irx3* mice. **a**, Energy expenditure over a 24-h period, corrected for lean mass (kcal kg^{-1} h^{-1}), for 18-week-old mice. **b**, Locomotor activity for mice in panel **a**. **c**, **d**, Average amount of food intake over a 24-h period with or without normalization to lean mass. **e**, Average locomotor activity measured over 24 h. **f**, **g**, Elevated gene and protein expression of Ucp1 in BAT of mutant mice (control/mutant, $n = 5/7$). Data are expressed as mean ± s.e.m. $*P < 0.05$ compared to control group.

doi:10.1038/nature13146

Endothelial Notch activity promotes angiogenesis and osteogenesis in bone

Saravana K. Ramasamy[1]*, Anjali P. Kusumbe[1]*, Lin Wang[1] & Ralf H. Adams[1,2]

Blood vessel growth in the skeletal system and osteogenesis seem to be coupled, suggesting the existence of molecular crosstalk between endothelial and osteoblastic cells[1,2]. Understanding the nature of the mechanisms linking angiogenesis and bone formation should be of great relevance for improved fracture healing or prevention of bone mass loss. Here we show that vascular growth in bone involves a specialized, tissue-specific form of angiogenesis. Notch signalling promotes endothelial cell proliferation and vessel growth in postnatal long bone, which is the opposite of the well-established function of Notch and its ligand Dll4 in the endothelium of other organs and tumours[3,4]. Endothelial-cell-specific and inducible genetic disruption of Notch signalling in mice not only impaired bone vessel morphology and growth, but also led to reduced osteogenesis, shortening of long bones, chondrocyte defects, loss of trabeculae and decreased bone mass. On the basis of a series of genetic experiments, we conclude that skeletal defects in these mutants involved defective angiocrine release of Noggin from endothelial cells, which is positively regulated by Notch. Administration of recombinant Noggin, a secreted antagonist of bone morphogenetic proteins, restored bone growth and mineralization, chondrocyte maturation, the formation of trabeculae and osteoprogenitor numbers in endothelial-cell-specific Notch pathway mutants. These findings establish a molecular framework coupling angiogenesis, angiocrine signals and osteogenesis, which may prove significant for the development of future therapeutic applications.

Dye injection and corrosion cast experiments had indicated that metaphyseal vessels in growing long bone are arranged in parallel, columnar strands with distal anastomotic interconnections[5,6]. Improved bone processing and immunostaining (see Methods) allowed the detailed analysis of these vessels, which were straight tubules with a diameter of 5–10 μm (Fig. 1a and Extended Data Fig. 1). At their distal end, next to growth plate chondrocytes, columns were linked by tubular arches, which carried blind-ended, bulb-shaped protrusions with a calibre of 15–20 μm (Fig. 1a and Extended Data Fig. 2a–c). Osterix^{+} osteoblasts and their progenitors were abundant around endothelial columns (Fig. 1a). Arch vessels extended filopodia towards the growth plate and were fully lumenized as demonstrated by Dextran perfusion (Fig. 1b), expression of lumen markers like ICAM2 or Podocalyxin, and electron microscopy (Extended Data Fig. 2d–f). Thus, the most distal, leading endothelial structures in long bone were not sprouts but tubular arches carrying bulb-like protrusions.

5-ethynyl-2′-deoxyuridine (EdU) labelling revealed that endothelial cell (EC) proliferation predominantly occurred within metaphyseal columns and, less prominently, in arches (Fig. 1c and Extended Data Fig. 2g). Endomucin (Emcn), a sialoprotein expressed by non-arterial endothelial cells[7], allowed the identification of arterioles in long bone as thin (5 μm) CD31^{+}Emcn^{-} vessels. Arterioles terminated in CD31^{+} Emcn^{+} vascular arches (Fig. 1d). As shown in the accompanying manuscript[8], CD31$^{\mathrm{hi}}$Emcn$^{\mathrm{hi}}$ (referred to here as type H) endothelial cells, as found in vessel columns and arches, display high proliferation rates and mediate vascular growth in bone.

Known regulators of vessel growth were highly expressed in freshly isolated type H endothelial cells relative to the much more abundant CD31$^{\mathrm{lo}}$Emcn$^{\mathrm{lo}}$ (type L) cells of the sinusoidal vasculature of the marrow cavity (Extended Data Fig. 3a). Immunostaining confirmed high expression of Neuropilin 1, Plexin D1, Unc5b or FLT4/VEGFR3 in vessel arches and columns (Extended Data Fig. 3b–g). These data indicate a specialized mode of angiogenesis in bone in which type H vessel arches and columns are the functional equivalent of endothelial sprouts.

Notch negatively regulates angiogenesis in embryos, murine retinas or tumours[3,4,9]. Notch and Delta-like 4 (Dll4), which signal in a cell–cell contact-dependent fashion between endothelial cells, limit sprouting and mitosis in the growing vasculature. In long bone, Dll4 was highly expressed by arteries and vessel arches. Jagged1, another Notch ligand controlling angiogenesis, was found in vessel arches and columns, but also in perivascular mesenchymal (osteoprogenitor) cells (Extended Data Fig. 3h, i). Inducible inactivation of the *Rbpj* gene encoding RBP-Jκ, an essential mediator of Notch-induced gene transcription, in postnatal endothelium unexpectedly decreased CD31$^{\mathrm{hi}}$Emcn$^{\mathrm{hi}}$ vessels and the number of total and proliferating endothelial cells in long bone (Fig. 1e–i and Extended Data Fig. 4a). The column/arch pattern was disrupted in *Rbpj*$^{\mathrm{i\Delta EC}}$ tibiae, and filopodia extension from mutant vessels was disorganized (Fig. 1e, g and Extended Data Fig. 4c–e). For Notch gain-of-function experiments, we targeted the gene for *Fbxw7*, which mediates polyubiquitination and proteasomal degradation of active Notch[10], with an inducible knockout strategy. Although mutants had strongly reduced retinal angiogenesis due to Notch overactivation[11], column/arch structures and both total and EdU^{+} endothelial cells were more abundant in *Fbxw7*$^{\mathrm{i\Delta EC}}$ long bone (Fig. 1f–i and Extended Data Fig. 4b, c). CD31$^{\mathrm{hi}}$ Emcn$^{\mathrm{hi}}$ vessel abundance was increased in *Fbxw7*$^{\mathrm{i\Delta EC}}$ and strongly decreased in *Rbpj*$^{\mathrm{i\Delta EC}}$ bone sections (Extended Data Fig. 4f–h). As Notch can positively regulate artery formation, *Fbxw7*$^{\mathrm{i\Delta EC}}$ bones contained numerous small arterioles (Extended Data Fig. 5a).

Suppression of angiogenesis by Dll4/Notch has been linked to VEGF receptor downregulation[3,4,9]. In *Rbpj*$^{\mathrm{i\Delta EC}}$ metaphyseal vessels, VEGFR1 (also known as FLT1), VEGFR2 (also known as KDR) and VEGFR3 (also known as FLT4) immunostaining was reduced, whereas *Fbxw7*$^{\mathrm{i\Delta EC}}$ samples contained more labelled vessels (Extended Data Fig. 5b–d). Transcripts for all three receptors were reduced in freshly sorted *Rbpj*$^{\mathrm{i\Delta EC}}$ bone endothelial cells but increased in *Fbxw7*$^{\mathrm{i\Delta EC}}$ cells (Fig. 1j and Extended Data Fig. 5e, f). In contrast, mRNA for soluble VEGFR1 (sFlt1), which antagonises VEGF-induced signalling, was strongly elevated in *Rbpj*$^{\mathrm{i\Delta EC}}$ samples and decreased in *Fbxw7*$^{\mathrm{i\Delta EC}}$ endothelial cells (Fig. 1k).

Inducible, endothelial-cell-specific overexpression of the Notch1 intracellular domain (NICD) phenocopied the *Fbxw7*$^{\mathrm{i\Delta EC}}$ results. CD31$^{\mathrm{hi}}$ Emcn$^{\mathrm{hi}}$ vessels and small arterioles were significantly more abundant in the mutant metaphysis (Extended Data Fig. 5g, h). Thus, although Notch inhibits vessel growth and endothelial proliferation in other organs, the pathway promotes bone angiogenesis. Expression of Notch target genes (*Hey1*, *Hes5*), modulators of angiogenesis (*Jag1*, *Ephb4*, *Angpt2*) and cell cycle regulators (*Cdkn1a*, *Cdkn1b*, *Cdkn2a*) were also

[1]Max Planck Institute for Molecular Biomedicine, Department of Tissue Morphogenesis, D-48149 Münster, Germany. [2]University of Münster, Faculty of Medicine, D-48149 Münster, Germany.
*These authors contributed equally to this work.

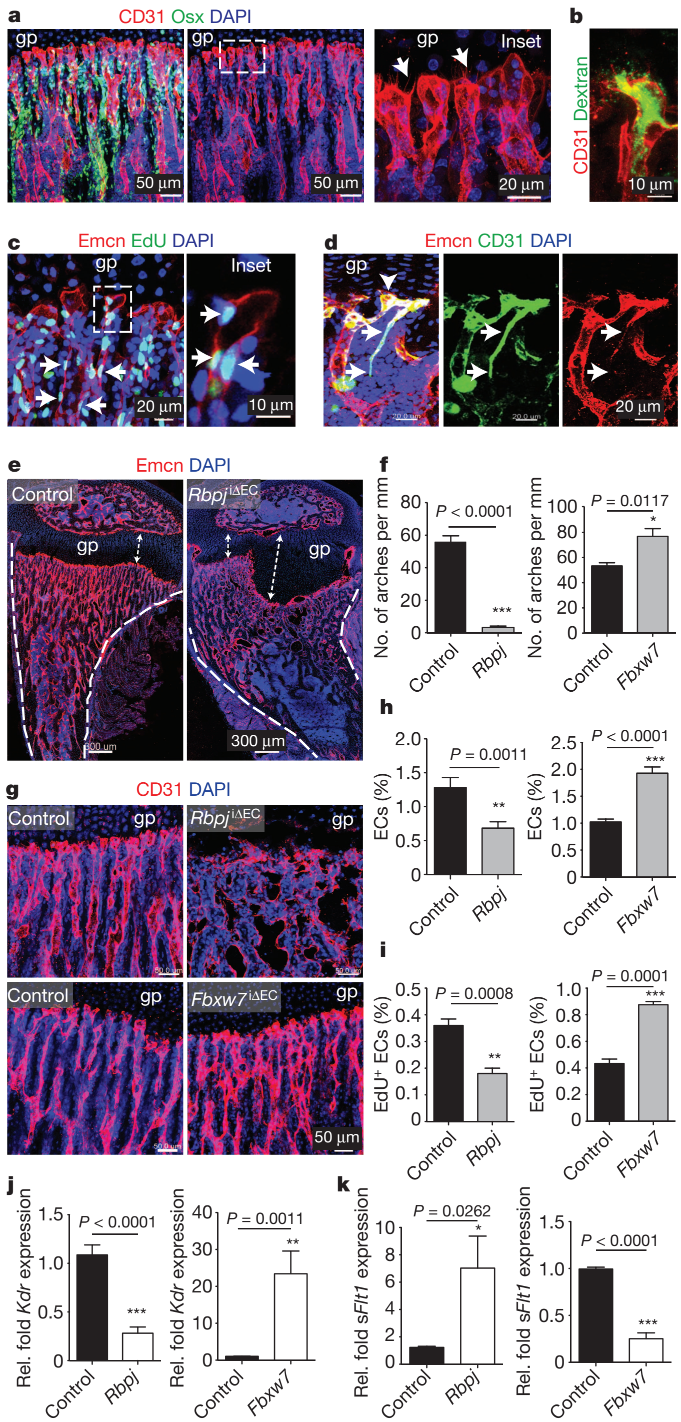


Figure 1 | Bone angiogenesis and regulation by Notch. **a**, Confocal images showing the metaphyseal vasculature in 4-week-old tibia. Osterix$^+$ (Osx, green) osteoblastic cells are closely associated with CD31$^+$ (red) endothelial cells. Nuclei, 4′,6-diamidino-2-phenylindole (DAPI, blue). Note parallel arrangement of vessels pointing towards growth plate (gp). Right panel shows higher magnification of inset with bulb-shaped endothelial protrusions and filopodia (arrows). **b**, Dextran perfusion of lumen in distal vessel within metaphysis of 4-week-old tibia. **c**, Maximum intensity projection showing proliferating (EdU$^+$, green) cells in metaphysis. Endothelial cells, Emcn (red); nuclei, DAPI (blue). Note abundant EdU$^+$ endothelial cells in columns. Inset shows single plane of EdU$^+$ arch endothelial cells. **d**, 10-week-old tibia showing connection between CD31$^+$Emcn$^-$ arteriole (arrows) and CD31$^+$Emcn$^+$ vascular arch (arrowhead). **e**, Defective metaphyseal vasculature and enlarged growth plate (gp, arrows) in *Rbpj*$^{i\Delta EC}$ tibia. Dashed lines, cortical bone; nuclei, DAPI (blue). **f**, Distal arches (per mm length of metaphysis) were reduced in *Rbpj*$^{i\Delta EC}$ tibiae and increased in *Fbxw7*$^{i\Delta EC}$ samples. ($n = 6$ mice bones from 4 independent litters), error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-tests. **g**, Confocal images showing defective CD31$^+$ vessels (red) in *Rbpj*$^{i\Delta EC}$ tibia and increased *Fbxw7*$^{i\Delta EC}$ vessel density. Nuclei, DAPI (blue). **h**, **i**, Flow cytometric quantitation of percentage of total (CD45$^-$Ter119$^-$ CD31$^+$) endothelial cells (**h**) ($n = 6$ mice from 4 independent litters) and EdU$^+$ endothelial cells (**i**) ($n = 5$ mice from 4 independent litters) in *Rbpj*$^{i\Delta EC}$ or *Fbxw7*$^{i\Delta EC}$ bone samples relative to controls. Error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-tests. **j**, **k**, qPCR analysis of VEGFR2 (*Kdr*) and soluble VEGFR1 (s*Flt1*) in sorted *Rbpj*$^{i\Delta EC}$ and *Fbxw7*$^{i\Delta EC}$ bone endothelial cells ($n = 4$ mice from 4 independent litters). Error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-test.

different between freshly isolated lung and bone endothelial cells (Extended Data Fig. 6a–d).

Defective angiogenesis in *Rbpj*$^{i\Delta EC}$ mice was accompanied by shortened long bones (Fig. 2a and Extended Data Fig. 7a) and disorganization of the metaphysis with large, irregular lacunae (Fig. 2b). Osteopontin immunostaining showed that mutants lacked clearly separated trabeculae and contained numerous small, highly interconnected bony elements (Fig. 2c). Micro-computed tomography (µ-CT) and histomorphometric analysis showed significant loss of bone mass and density (that is, bone volume over total volume; BV/TV), number and thickness of trabeculae, and osteoid thickness in *Rbpj*$^{i\Delta EC}$ tibiae, without appreciable changes in bone-resorbing osteoclasts (Fig. 2d, e and Extended Data Fig. 7b–j). Calcein double labelling showed strongly reduced bone formation rates in *Rbpj*$^{i\Delta EC}$ bone (Fig. 2f, g). Endothelial-specific Notch activation in *Fbxw7*$^{i\Delta EC}$ mutants had the opposite effect (Extended Data Fig. 7m–t).

Osterix$^+$ cell numbers were significantly increased in the *Rbpj*$^{i\Delta EC}$ metaphysis (Fig. 2h, i). Conversely, stabilization of active Notch in *Fbxw7*$^{i\Delta EC}$ mice significantly enhanced femur extension and bone growth and reduced Osterix$^+$ cell abundance without appreciable effect on osteoclasts (Fig. 2j and Extended Data Figs 7j, k and 8a). *Bglap* (bone gamma-carboxyglutamate protein) and *Ibsp* (integrin-binding sialoprotein) expression were reduced in *Rbpj*$^{i\Delta EC}$ bones (Fig. 2k) indicating that accumulating Osterix$^+$ cells were immature osteoblasts. Runx2$^+$ early osteoprogenitors, which give rise to Osterix$^+$ osteoblasts[12], were strongly reduced in *Rbpj*$^{i\Delta EC}$ samples (Fig. 2l, m) but more abundant in *Fbxw7*$^{i\Delta EC}$ mutants (Fig. 2n and Extended Data Fig. 8a). Confirming the accumulation of osteoprogenitors, primary mesenchymal cells from *Rbpj*$^{i\Delta EC}$ bone generated mineral nodules within 10–14 days, whereas controls required 21–28 days (Extended Data Fig. 8b and data not shown). These data suggested that Notch-dependent, endothelial-cell-derived signals control osteoprogenitor differentiation. Arguing against defective oxygenation as the cause of these changes, pimonidazole and anti-HIF1α staining were not appreciably altered in *Rbpj*$^{i\Delta EC}$ long bone (Extended Data Fig. 8c, d).

To identify the relevant Notch ligand in bone vessels, *Cdh5(PAC)-CreERT2* transgenics with endothelial-cell-specific expression of tamoxifen-inducible Cre recombinase were combined with conditional *Dll1* (ref. 13), *Dll4* (ref. 14) or *Jag1* (encoding Jagged1)[15] mice, respectively. Vessel growth and patterning was not visibly altered in the *Dll1*$^{i\Delta EC}$ or *Jag1*$^{i\Delta EC}$ mutant skeleton (Extended Data Fig. 8e). *Dll4*$^{i\Delta EC}$ long bones were substantially shorter and phenocopied *Rbpj*$^{i\Delta EC}$ defects (Extended Data Fig. 8f, g) including loss of trabeculae, appearance of highly interconnected bony elements, accumulation of Osterix$^+$ cells and reduction of Runx2$^+$ early osteoprogenitors (Fig. 3a, g and data not shown). Highly similar *Dll4*$^{i\Delta EC}$ and *Rbpj*$^{i\Delta EC}$ phenotypes argued that RBP-Jκ is essential for Notch signalling and not primarily acting as a transcriptional repressor[16].

Rbpj$^{i\Delta EC}$ and *Dll4*$^{i\Delta EC}$ mutants displayed enlarged and irregularly shaped growth plates (Figs 1e, 2c and 3a). While labelling of mitotic chondrocytes did not reveal alterations in the proliferating zone, mutant maturation/hypertrophy zones were significantly enlarged and expression of Sox9, which regulates chondrocyte differentiation and hypertrophy, was lost (Fig. 3c–e and Extended Data Fig. 8h, i). VEGF-A expression by mature and hypertrophic chondrocytes[17] was also strongly reduced

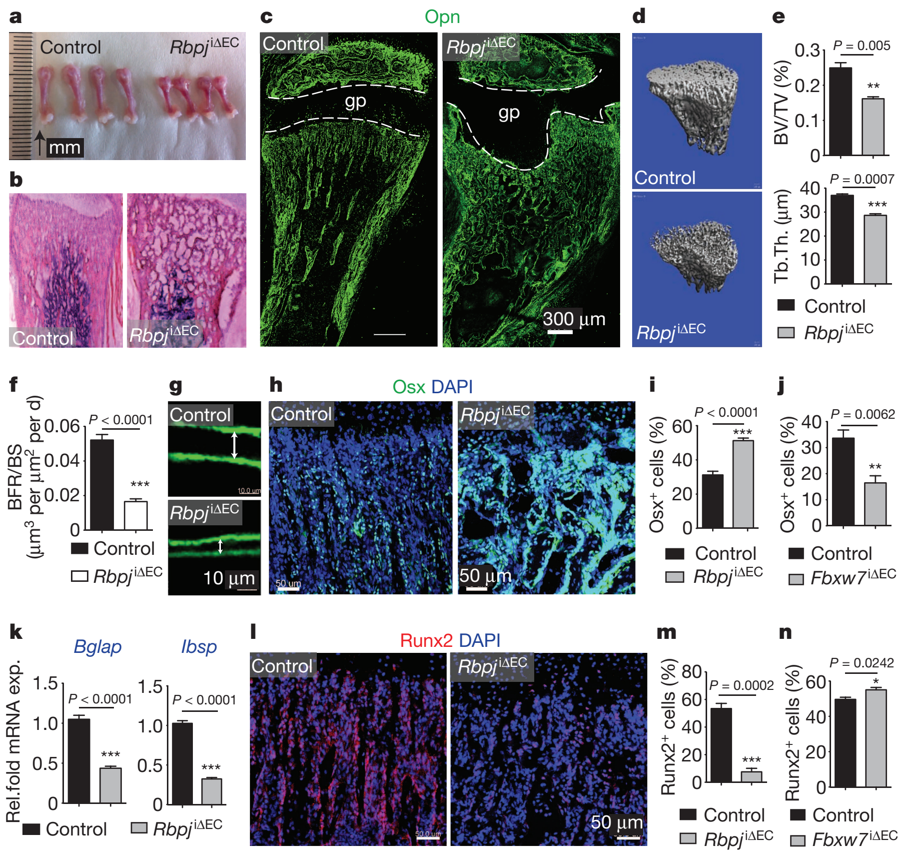


Figure 2 | Endothelial Notch signalling regulates osteogenesis. a, Decreased length of freshly dissected *Rbpj*$^{i\Delta EC}$ femurs compared to control littermates. **b**, Haematoxylin & eosin (H&E) stained longitudinal *Rbpj*$^{i\Delta EC}$ and control tibia sections. **c**, Osteopontin (Opn) immunostaining showing defective formation of trabeculae in P28 *Rbpj*$^{i\Delta EC}$ tibia. **d**, Three-dimensional μ-CT reconstruction of 4-week-old control and *Rbpj*$^{i\Delta EC}$ metaphysis. **e**, Reduced trabecular bone volume density measured as bone volume/total volume (BV/TV) and trabecular bone thickness (Tb.Th) in *Rbpj*$^{i\Delta EC}$ mice. **f**, **g**, Bone formation rate per bone surface (BFR/BS, **f**) calculated by calcein double labelling (7 day time interval) confirmed decreased bone formation in P28 *Rbpj* mutants. Arrows in **g** mark distance between calcein-labelled layers ($n = 10$ mice from 6 independent litters). Error bars, ± s.e.m. *P* value, two-tailed unpaired *t*-tests. **h**, Osterix (Osx) immunostaining shows strongly increased osteoprogenitor numbers in *Rbpj*$^{i\Delta EC}$ metaphysis. **i**, **j**, Quantitation of metaphyseal Osx$^+$ cells in *Rbpj*$^{i\Delta EC}$ (**i**) and *Fbxw7*$^{i\Delta EC}$ mutants (**j**) relative to controls ($n = 6$ mice from 4 independent litters). Error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-tests. **k**, qPCR analysis showing reduced expression of mature osteoblast markers (*Bglap*, *Ibsp*) in *Rbpj*$^{i\Delta EC}$ bones ($n = 6$ mice from 4 independent litters). **l**, Immunostaining showing decrease of Runx2$^+$ early osteoprogenitors in 4-week-old *Rbpj*$^{i\Delta EC}$ tibiae. **m**, **n**, Quantitation of metaphyseal Runx2$^+$ cells in *Rbpj*$^{i\Delta EC}$ (**m**) and *Fbxw7*$^{i\Delta EC}$ mutants (**n**) relative to littermate controls ($n = 6$ mice from 4 independent litters). Error bars ± s.e.m. *P* values, two-tailed unpaired *t*-tests.

(Fig. 3f and Extended Data Fig. 8j, k). The size of *Fbxw7*$^{i\Delta EC}$ growth plates and maturation/hypertrophy zones was reduced, whereas Sox9 expression was maintained (Extended Data Fig. 9a–c, f, g). *Vegfa* transcripts were elevated in *Fbxw7*$^{i\Delta EC}$ bone samples in sharp contrast to the reduction in *Rbpj*$^{i\Delta EC}$ tibiae (Extended Data Fig. 9d, e).

Mesenchymal Notch expression promotes proliferation and suppresses osteoblast differentiation[18,19]. As *Dll4* expression is transcriptionally regulated by Notch and RBP-Jκ[3], endothelial Dll4 might signal through Notch receptors on perivascular osteoprogenitors. However, targeting of *Rbpj* in postnatal osteoblastic cells with *Tg(Col1a1-creERT2) 6.1.ICS* transgenic mice (Extended Data Fig. 9h–j) with the same experimental schedule as for endothelial-cell-specific mutants did not lead to appreciable defects in the metaphyseal vasculature or osteoprogenitor cells (Fig. 3b and Extended Data Fig. 9k and data not shown). Reinstating Notch signalling in endothelial cells by combining *Dll4*$^{i\Delta EC}$ and NICD overexpressing mutants (*NICD*$^{iOE\text{-}EC}$) rescued the structural alterations seen in the *Dll4*$^{i\Delta EC}$ metaphysis and bone vasculature, enabled trabeculae formation and restored growth plate size, Sox9 expression and osteoprogenitor numbers (Fig. 3g and Extended Data Fig. 9l–p). Thus, endothelial-cell-specific and cell-autonomous Notch activity regulates bone angiogenesis and couples it to osteogenesis.

Recent studies have highlighted the importance of angiocrine factors for organ growth and regeneration[20]. To identify potential endothelial-cell-derived signals acting on osteoblastic cells, endothelial cells isolated from *Fbxw7*$^{i\Delta EC}$ and control bone were screened by a candidate approach. *Tgfb2*, *Bmp4*, *Fgf1*, *Fgf8*, *Wnt1*, *Wnt3a*, *Wnt10b*, *Dkk1* and *Pgf* (encoding Plgf) transcripts were increased in *Fbxw7*$^{i\Delta EC}$ bone endothelium (Fig. 4a). The highest upregulation (40-fold) was observed for *Nog* mRNA encoding Noggin, a secreted antagonist of bone morphogenetic proteins (BMPs). Other secreted BMP antagonists, namely Chordin (*Chrd*) and Cerberus 1 (*Cer1*), were not detectable in FACS-isolated bone endothelial cells (Fig. 4a). As additional evidence for Noggin as a Notch-controlled angiocrine regulator, *Nog* expression was significantly reduced in purified *Rbpj*$^{i\Delta EC}$ bone endothelial cells (Fig. 4b). In control tibiae, Noggin immunostaining was detected in endothelial columns/arches and surrounding perivascular cells. In diaphysis, Noggin was detectable at lower level and without pronounced vessel association (Extended Data Fig. 10a, b). Consistent with its known role in bone formation[21–23], Noggin inhibited primary mesenchymal cell differentiation into osteoblasts *in vitro* (Fig. 4c). Whereas global *Nog* gene knockouts were embryonic lethal, heterozygotes and osteoblast-specific mutants developed skeletal defects reminiscent of *Rbpj*$^{i\Delta EC}$ mice including impaired bone

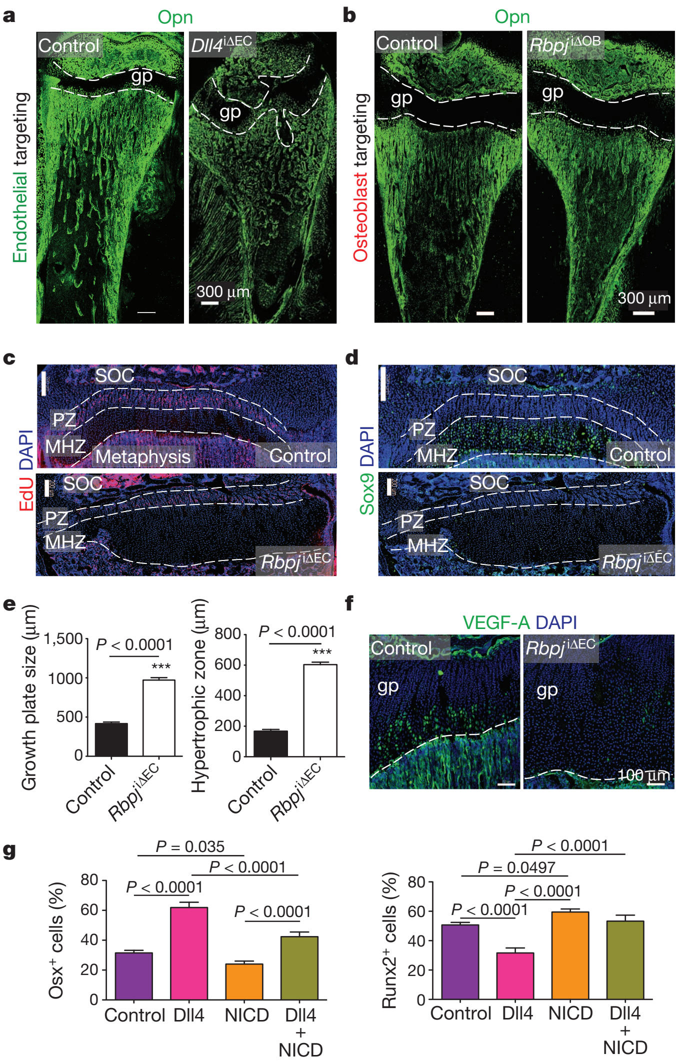


Figure 3 | Osteogenesis requires EC-autonomous Notch signalling. **a**, **b**, Osteopontin (Opn) immunostaining showing malformed bone and growth plate (gp, dashed lines) in 4-week-old $Dll4^{i\Delta EC}$ (**a**) but not in osteoblast-specific *Rbpj* mutant ($Rbpj^{iOB}$) tibiae (**b**). **c**, **d**, Confocal images of control and $Rbpj^{i\Delta EC}$ EdU-labelled proliferating zones (PZ, **c**) and Sox9 immunostained maturation/hypertrophy zones (MHZ, **d**). Note diminished Sox9 in $Rbpj^{i\Delta EC}$ mutant. Dashed lines mark PZ and MHZ borders. SOC, secondary ossification centre; nuclei, DAPI (blue). **e**, Quantitative analysis of the sizes of $Rbpj^{i\Delta EC}$ mutant growth plates and maturation/hypertrophy zones ($n = 6$ mice from 4 independent litters). Error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-test. **f**, Strongly reduced VEGF-A immunostaining (green) in P28 $Rbpj^{i\Delta EC}$ mutant growth plate (gp) chondrocytes. **g**, Quantitative analysis of Osx^+ (left) and $Runx2^+$ (right) cells in control, *Dll4*, *NICD* or $Dll4^+$ *NICD* tibiae, as indicated. Controls are littermate animals without Cre expression ($n = 4$ mice from 4 independent litters). Error bars, ± s.e.m. *P* values, one-way ANOVA with Bonferroni's multiple comparison post-hoc test.

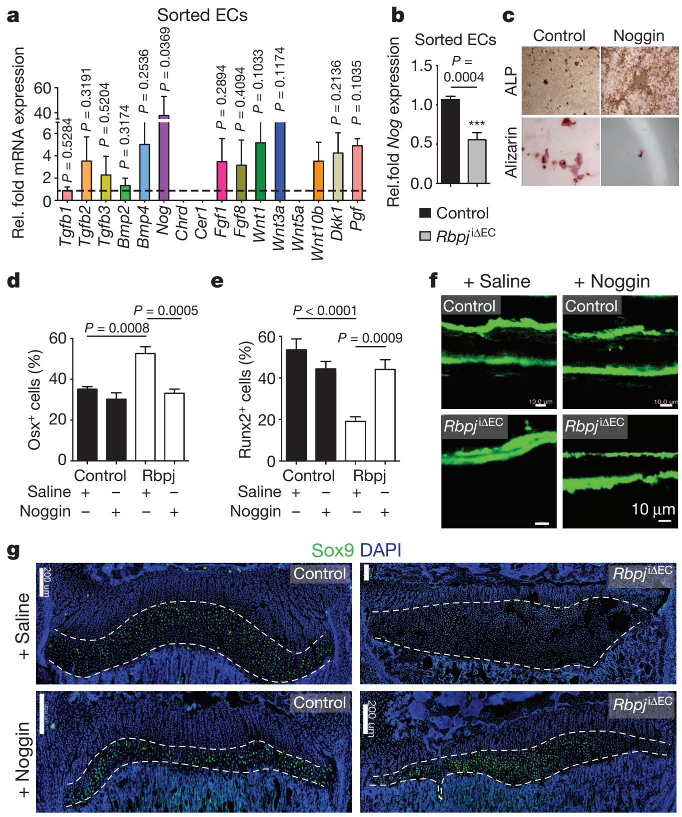


Figure 4 | Role of Notch-dependent, angiocrine Noggin expression. **a**, *Tgfb1*, *Tgfb2*, *Tgfb3*, *Bmp2*, *Bmp4*, *Nog*, *Chrd*, *Cer1*, *Fgf1*, *Fgf8*, *Wnt1*, *Wnt3a*, *Wnt5a*, *Wnt10b*, *Dkk1* and *Pgf* mRNA expression in isolated bone endothelial cells from $Fbxw7^{i\Delta EC}$ mice normalized to littermate controls (dashed line) ($n = 4$ mice from 4 independent litters). Error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-test. **b**, qPCR analysis showing reduced *Nog* expression in endothelial cells sorted from $Rbpj^{i\Delta EC}$ long bone mice ($n = 4$ mice from 4 independent litters). Error bars, ± s.e.m. *P* value, two-tailed unpaired *t*-test. **c**, Noggin-mediated inhibition of osteoblastic differentiation of cultured murine mesenchymal progenitors. Mineral nodule formation (Alizarin staining) was supressed, and alkaline phosphatase (ALP) retained by Noggin after 28 days of *in vitro* differentiation. **d**, **e**, Quantitation of Osx^+ (**d**) and $Runx2^+$ cells (**e**) in Noggin-treated versus saline control and $Rbpj^{i\Delta EC}$ long bones ($n = 6$ mice from 4 independent litters). Error bars, ± s.e.m. *P* values, one-way ANOVA with Bonferroni's multiple comparison post-hoc test. **f**, Calcein double labelling in $Rbpj^{i\Delta EC}$ mutant or control tibiae treated with saline or Noggin, as indicated. Noggin strongly improved $Rbpj^{i\Delta EC}$ mineral apposition rates. **g**, Noggin restored $Rbpj^{i\Delta EC}$ growth plate size and Sox9 expression (green) in the maturation/hypertrophy zone (marked by dashed lines). Nuclei, DAPI (blue).

formation, shortened femoral length, reduced trabeculae and oversized growth plates[22,23]. Daily injection of $Rbpj^{i\Delta EC}$ mice with recombinant Noggin protein for 2 weeks before analysis at P28 restored metaphysis organization, improved bone formation and normalized the number of osteoprogenitors (Fig. 4d–f and Extended Data Fig. 10c–f). Exogenous Noggin also restored the organization of the $Rbpj^{i\Delta EC}$ bone vasculature, growth plate size, and chondrocyte Sox9 and VEGF-A expression (Fig. 4g and Extended Data Fig. 10g–i).

These findings identify Notch as an important positive regulator of vascular growth in bone, which is highly distinct from the Dll4/Notch-mediated suppression of endothelial cell mitosis and sprouting in other organs. Whereas previous work had already indicated that $Osterix^+$ cells co-invade with growing vessels in experimental bone fractures[2], we now establish a molecular framework coupling the activity of endothelial cells, chondrocytes and osteoprogenitors. Future work will address whether Notch and other pathways can be used to promote fracture healing or prevent pathological bone loss.

METHODS SUMMARY

For wild-type bone analysis, C57BL/6J male mice were used unless stated otherwise. All endothelial-cell-specific gene targeting experiments were performed with *Cdh5(PAC)-CreERT2* transgenic mice[24]. For *Rbpj* deletion in the postnatal endothelium, mice carrying *loxP*-flanked *Rbpj* ($Rbpj^{lox/lox}$) alleles[25] and *Cdh5(PAC)-CreERT2* transgenics were interbred. To induce Cre activity and gene deletion, offspring was injected with 500 µg tamoxifen (Sigma, T5648) intraperitoneally every day from P10 to P14. The resulting $Rbpj^{i\Delta EC}$ ($CreERT2^{T/+}$ $Rbpj^{lox/lox}$) mutants and Cre- littermate controls were sacrificed at P28, and femurs and tibiae were collected for analysis. Identical breeding and tamoxifen administration strategies were used to generate EC-specific mutants with $Fbxw7^{lox/lox}$ (ref. 10), $Dll1^{lox/lox}$

(ref. 13), *Dll4*$^{lox/lox}$ (ref. 14), or *Jag1*$^{lox/lox}$ (ref. 15) mice and inducible osteoblast-specific *Rbpj* knockouts (*Rbpj*^{iOB}). The latter were generated with *Tg(Col1a1-CreERT2)6.1.ICS* transgenic mice (Institut Clinique de la Souris, France).

For Notch gain-of-function experiments, *Gt(ROSA)26Sor*$^{tm1(Notch1)Dam/J}$ mice carrying a Cre-inducible transgene for Notch1 intracellular domain overexpression[26] and *Cdh5(PAC)-CreERT2* transgenics were interbred. Tamoxifen administration (see above for injection schedule) was used to generate *CreERT2*-positive (NICD$^{iOE\text{-}EC}$) mutants overexpressing NICD in endothelial cells and controls. For experiments with endothelial-cell-specific *Dll4*$^{i\Delta EC}$ NICD$^{iOE\text{-}EC}$ double mutants, interbreeding of *Dll4*$^{lox/lox}$ conditional mice[14] with *Gt(ROSA)26Sor*$^{tm1(Notch1)Dam/J}$ and *Cdh5(PAC)-CreERT2* transgenics generated triple heterozygote males, which were bred to *Dll4*$^{lox/+}$ *Gt(ROSA)26Sor*$^{tm1(Notch1)Dam/J}$ double heterozygous females. This produced *CreERT2*-negative controls together with *Dll4*$^{lox/lox}$ NICD$^{+/+}$ *CreERT2*$^{T/+}$ (*Dll4*$^{i\Delta EC}$) and *Dll4*$^{+/+}$ NICD$^{+/OE}$ *CreERT2*$^{T/+}$ (NICD$^{iOE\text{-}EC}$) single mutants, and *Dll4*$^{i\Delta EC}$/NICD$^{iOE\text{-}EC}$ double mutants, which received tamoxifen and were analysed as described above. For mutant analysis, both male and female mice were used.

For Noggin administration experiments, mice were injected intraperitoneally once daily with 500 μg per kg recombinant Noggin (R&D Systems) from P15 to P27, after completion of tamoxifen injections (P10–P14) and before analysis at P28.

All animal experiments were performed in compliance with the relevant laws and institutional guidelines and were approved by local animal ethics committees.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 July 2013; accepted 11 February 2014.

Published online 12 March; corrected online 19 March 2014 (see full-text HTML version for details).

1. Eshkar-Oren, I. *et al.* The forming limb skeleton serves as a signaling center for limb vasculature patterning via regulation of *Vegf*. *Development* **136,** 1263–1272 (2009).
2. Maes, C. *et al.* Osteoblast precursors, but not mature osteoblasts, move into developing and fractured bones along with invading blood vessels. *Dev. Cell* **19,** 329–344 (2010).
3. Benedito, R. *et al.* The notch ligands Dll4 and Jagged1 have opposing effects on angiogenesis. *Cell* **137,** 1124–1135 (2009).
4. Noguera-Troise, I. *et al.* Blockade of Dll4 inhibits tumour growth by promoting non-productive angiogenesis. *Nature* **444,** 1032–1037 (2006).
5. Skawina, A., Litwin, J. A., Gorczyca, J. & Miodonski, A. J. The vascular system of human fetal long bones: a scanning electron microscope study of corrosion casts. *J. Anat.* **185,** 369–376 (1994).
6. Trueta, J. & Morgan, J. D. The vascular contribution to osteogenesis. I. Studies by the injection method. *J. Bone Joint Surg. Br.* **42-B,** 97–109 (1960).
7. Kuhn, A. *et al.* Expression of endomucin, a novel endothelial sialomucin, in normal and diseased human skin. *J. Invest. Dermatol.* **119,** 1388–1393 (2002).
8. Kusumbe, A. P., Ramasamy, S. K. & Adams, R. H. Coupling of angiogenesis and osteogenesis by a specific vessel subtype in bone. *Nature* http://dx.doi.org/10.1038/nature13145 (this issue).
9. Siekmann, A. F. & Lawson, N. D. Notch signalling limits angiogenic cell behaviour in developing zebrafish arteries. *Nature* **445,** 781–784 (2007).
10. Hoeck, J. D. *et al.* Fbw7 controls neural stem cell differentiation and progenitor apoptosis via Notch and c-Jun. *Nature Neurosci.* **13,** 1365–1372 (2010).
11. Izumi, N. *et al.* Fbxw7 controls angiogenesis by regulating endothelial Notch activity. *PLoS ONE* **7,** e41116 (2012).
12. Nakashima, K. *et al.* The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. *Cell* **108,** 17–29 (2002).
13. Hozumi, K. *et al.* Delta-like 1 is necessary for the generation of marginal zone B cells but not T cells *in vivo*. *Nature Immunol.* **5,** 638–644 (2004).
14. Koch, U. *et al.* Delta-like 4 is the essential, nonredundant ligand for Notch1 during thymic T cell lineage commitment. *J. Exp. Med.* **205,** 2515–2523 (2008).
15. Brooker, R., Hozumi, K. & Lewis, J. Notch ligands with contrasting functions: Jagged1 and Delta1 in the mouse inner ear. *Development* **133,** 1277–1286 (2006).
16. Olave, I., Reinberg, D. & Vales, L. D. The mammalian transcriptional repressor RBP (CBF1) targets TFIID and TFIIA to prevent activated transcription. *Genes Dev.* **12,** 1621–1637 (1998).
17. Gerber, H. P. *et al.* VEGF couples hypertrophic cartilage remodeling, ossification and angiogenesis during endochondral bone formation. *Nature Med.* **5,** 623–628 (1999).
18. Engin, F. *et al.* Dimorphic effects of Notch signaling in bone homeostasis. *Nature Med.* **14,** 299–305 (2008).
19. Hilton, M. J. *et al.* Notch signaling maintains bone marrow mesenchymal progenitors by suppressing osteoblast differentiation. *Nature Med.* **14,** 306–314 (2008).
20. Potente, M., Gerhardt, H. & Carmeliet, P. Basic and therapeutic aspects of angiogenesis. *Cell* **146,** 873–887 (2011).
21. Wu, X. B. *et al.* Impaired osteoblastic differentiation, reduced bone formation, and severe osteoporosis in noggin-overexpressing mice. *J. Clin. Invest.* **112,** 924–934 (2003).
22. Canalis, E., Brunet, L. J., Parker, K. & Zanotti, S. Conditional inactivation of noggin in the postnatal skeleton causes osteopenia. *Endocrinology* **153,** 1616–1626 (2012).
23. Tylzanowski, P., Mebis, L. & Luyten, F. P. The Noggin null mouse phenotype is strain dependent and haploinsufficiency leads to skeletal defects. *Dev. Dyn.* **235,** 1599–1607 (2006).
24. Wang, Y. *et al.* Ephrin-B2 controls VEGF-induced angiogenesis and lymphangiogenesis. *Nature* **465,** 483–486 (2010).
25. Han, H. *et al.* Inducible gene knockout of transcription factor recombination signal binding protein-J reveals its essential role in T versus B lineage decision. *Int. Immunol.* **14,** 637–645 (2002).
26. Murtaugh, L. C., Stanger, B. Z., Kwan, K. M. & Melton, D. A. Notch signaling controls multiple steps of pancreatic differentiation. *Proc. Natl Acad. Sci. USA* **100,** 14920–14925 (2003).

Acknowledgements We thank A. Duarte, F. Radtke and T. Honjo for floxed *Dll4* and *Rbpj* mutant mice, A. Borgscheiper for technical assistance, M. Stehling for endothelial cell sorting, D. Zeuschner for electron microscopy. Funding was provided by the Max Planck Society, the University of Münster, the DFG cluster of excellence 'Cells in Motion', and the European Research Council (AdG 339409 AngioBone).

Author Contributions S.K.R., A.P.K and R.H.A. designed experiments and interpreted results. S.K.R. generated and characterized mouse mutant lines. S.K.R. and A.P.K. did all experiments including FACS, qRT-PCR, bone sectioning and staining, confocal imaging and quantifications. L.W. performed analysis of bone vessel ultrastructure. S.K.R. and R.H.A. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.H.A. (ralf.adams@mpi-muenster.mpg.de).

METHODS

Genetically modified mice. For wild-type bone analysis, C57BL/6J male mice were used unless stated otherwise. All endothelial-cell-specific gene targeting experiments were performed with *Cdh5(PAC)-CreERT2* transgenic mice[24]. For *Rbpj* deletion in the postnatal endothelium, mice carrying *loxP*-flanked *Rbpj* ($Rbpj^{lox/lox}$) alleles[25] and *Cdh5(PAC)-CreERT2* transgenics were interbred. To induce Cre activity and gene deletion, offspring was injected with 500 μg tamoxifen (Sigma, T5648) intraperitoneally every day from P10 to P14. The resulting $Rbpj^{i\Delta EC}$ ($CreERT2^{T/+}$ $Rbpj^{lox/lox}$) mutants and Cre- littermate controls were sacrificed at P28, and femurs and tibiae were collected for analysis. Identical breeding and tamoxifen administration strategies were used to generate EC-specific mutants with $Fbxw7^{lox/lox}$ (ref. 10), $Dll1^{lox/lox}$ (ref. 13), $Dll4^{lox/lox}$ (ref. 14), or $Jag1^{lox/lox}$ (ref. 15) mice and inducible osteoblast-specific *Rbpj* knockouts ($Rbpj^{iOB}$). The latter were generated with *Tg(Col1a1-creERT2)6.1.ICS* transgenic mice (Institut Clinique de la Souris, France).

For Notch gain-of-function experiments, $Gt(ROSA)26Sor^{tm1(Notch1)Dam/J}$ mice carrying a Cre-inducible transgene for Notch1 intracellular domain overexpression[26] and *Cdh5(PAC)-CreERT2* transgenics were interbred. Tamoxifen administration (see above for injection schedule) was used to generate *CreERT2*-positive ($NICD^{iOE\text{-}EC}$) mutants overexpressing NICD in endothelial cells and controls. For experiments with endothelial-cell-specific $Dll4^{i\Delta EC}$ $NICD^{iOE\text{-}EC}$ double mutants, interbreeding of $Dll4^{lox/lox}$ conditional mice[14] with $Gt(ROSA)26Sor^{tm1(Notch1)Dam/J}$ and *Cdh5(PAC)-CreERT2* transgenics generated triple heterozygote males, which were bred to $Dll4^{lox/+}$ $Gt(ROSA)26Sor^{tm1(Notch1)Dam/J}$ double heterozygous females. This produced *CreERT2*-negative controls together with $Dll4^{lox/lox}$ $NICD^{+/+}$ $CreERT2^{T/+}$ ($Dll4^{i\Delta EC}$) and $Dll4^{+/+}$ $NICD^{+/OE}$ $CreERT2^{T/+}$ ($NICD^{iOE\text{-}EC}$) single mutants, and $Dll4^{i\Delta EC}$/$NICD^{iOE\text{-}EC}$ double mutants, which received tamoxifen and were analysed as described above. For mutant analysis, both male and female mice were used.

For Noggin administration experiments, mice were injected intraperitoneally once daily with 500 μg per kg recombinant Noggin (R&D Systems) from P15 to P27, after completion of tamoxifen injections (P10–P14) and before analysis at P28.

All animal experiments were performed in compliance with the relevant laws and institutional guidelines and were approved by local animal ethics committees.

Immunohistochemistry. Freshly dissected tibiae were collected from wild-type and mutant mice and their control littermates and immediately fixed in 4% ice-cold paraformaldehyde solution for 4 h. Decalcification was carried out with 0.5 M EDTA at 4 °C with constant shaking for 24 h. Then the decalcified bones were immersed into 20% sucrose and 2% polyvinylpyrrolidone (PVP) for 24 h. Finally, the tissues were embedded and frozen in 8% gelatin (porcine) in the presence of 20% sucrose and 2% PVP and cryosectioned using low-profile blades on a Leica CM3050 cryostat.

For phenotypic analysis, mutant mice and littermate controls were always processed, sectioned, stained, imaged and analysed together at the same conditions and settings. For immunostaining, bone sections were air-dried, permeabilized for 10 min in 0.3% Triton X-100, blocked in 5% donkey serum at room temperature for 30 min. Blocked sections were probed with the primary antibodies diluted in 5% donkey serum in PBS for 2 h at room temperature or overnight at 4 °C. After primary antibody incubation, sections were washed with PBS for three times and incubated with appropriate Alexa Fluor-coupled secondary antibodies (1:400, Molecular Probes) for 1 h at room temperature. Nuclei were counterstained with DAPI. Sections were thoroughly washed with PBS before mounting them using FluoroMount-G (Southern Biotech). Finally, the slides were air-dried and sealed with nail polish.

The following primary antibodies were used: Endomucin (sc-65495, Santa Cruz, diluted 1:100), Pecam1 conjugated to Alexa Fluor488 (FAB3628G, R&D Systems, 1:100), Pecam1 (553370, BD Pharmingen, 1:100), Osterix (sc-22536-R, Santa Cruz, 1:200), Runx2 (MAB2006, R&D Systems, 1:200), α-SMA-Cy3 (C6198, Sigma, 1:100), Hif1-α (ab65979, Abcam, 1:100), Collagen type 1 (AB675P, Millipore, 1:200), Osteopontin (AF808, R&D Systems, 1:200), Calcitonin receptor (ab11042, Abcam, 1:75), Noggin (SC-25656, Santa Cruz, 1:200), Icam2 (553326, BD Pharmingen, 1:100), Podocalyxin (AF1556, R&D Systems, 1:100), green fluorescent protein (GFP) (A21311, Invitrogen, 1:200), VEGF-A (SC-152, Santa Cruz, 1:100), Sox9 (AF3075, R&D Systems, 1:100), Dll4 (AF1389, R&D Systems, 1:100), Jag1 (SC-6011, Santa Cruz, 1:200), VEGFR1 (NB100-57643, Novus Biologicals, 1:100), VEGFR2 (555307, BD Pharmingen, 1:100), VEGFR3 (AF743, R&D Systems, 1:100), Neuropilin 1 (ab81321, Abcam, 1:100), Neuropilin 2 (AB10522, Millipore, 1:200), Plexin D1 (SC-67145, Santa Cruz, 1:200), Unc5b (ab104871, Abcam, 1:200), Robo 4 (ab10547, Abcam, 1:100), Biotin-conjugated CD45 (553077, Becton Dickinson, 1:100), Biotin-conjugated Ter-119 (553672, Becton Dickinson, 1:100),

For visualizing vascular lumina in the angiogenic front, tetramethyl rhodamine-conjugated fixable dextran of 2,000 kDa (Invitrogen, D7139) was perfused through the heart of deeply anaesthetized mice. Mice were euthanized after 10 min and bones were processed further.

For labelling of proliferating cells, mice were intraperitoneally injected with 300 μg EdU (Invitrogen) 3 h before euthanasia. Tibiae were immediately collected and processed as mentioned earlier. The bone sections were immunostained for EDU using Click-iT chemistry following the manufacturer's protocols (Invitrogen).

For metabolically labelling with the hypoxia probe pimonidazole (Pimo, Hypoxyprobe Inc.), mutant and control mice were intraperitoneally injected with 60 mg per kg Pimo for 2 h before euthanasia. Metabolized Pimo was detected by a rabbit antiserum against the non-oxidized, protein-conjugated form of pimonidazole (Hypoxyprobe Inc.).

Image acquisition, processing and quantitative analysis. Immunofluorescent stainings were analysed at high resolution with a Zeiss LSM-780 laser scanning confocal microscope. *Z*-stacks of images were processed and 3D reconstructed with the Imaris software (version 7.00, Bitplane). Imaris, Image J, Photoshop and Illustrator (Adobe) softwares were used for image processing, in compliance with *Nature*'s guide for digital images.

For all quantifications, a region of 300–400 μm from growth plate towards the caudal region was selected for all cell number quantifications in the metaphysis. All quantifications were done with Imaris and Image J software on high-resolution confocal images.

Mesenchymal cell culture and differentiation. Primary mesenchymal cells were isolated from mouse bone and cultured following published protocols[27,28]. In brief, freshly isolated long bones were crushed and then digested with collagenase at 37 °C for 1 h. Cells and small bone chips were washed thoroughly and cultured in 2% FBS containing alpha-MEM for few days until adherent mesenchymal cells had migrated out from the bone fragments. The adherent cells were passaged further and maintained at 37 °C with 5% CO_2 in a humidified atmosphere. For osteogenic differentiation, StemXVivo osteogenic basal medium and its supplement (RnD systems, CCM009) were used. Continuous treatment of recombinant mouse Noggin (500 ng ml^{-1}, R&D systems) was provided to cells and, after 28 days, the appearance of mineral nodules and mesenchymal progenitors was analysed using Alizarin red and alkaline phosphatase staining, respectively.

Quantitative RT–PCR (qRT–PCR). For the analysis of mRNA expression levels, pure endothelial cell fractions were flow-sorted directly into the lysis buffer of RNeasy Mini Kit (Qiagen). Total RNA was isolated according to manufacturer's protocol. A total of 100–500 ng RNA per reaction was used to generate cDNA with the iScript cDNA Synthesis System (Bio-Rad). Quantitative PCR (qPCR) was performed using TaqMan gene expression assays on ABI PRISM 7900HT Sequence Detection System. The FAM-conjugated TaqMan probes *Kdr*, *Flt4*, membrane isoform *Flt1*, customised s*Flt1*, *Nrp1*, *Plexind1*, *Unc5b*, *Robo4*, *Ntn1*, *Ntn3*, *Dll4*, *Jag1*, *Efnb2*, *Ephb4*, *Hes1*, *Hey1*, *Hey2*, *Hes5*, *Icam1*, *Angpt2*, *Cdh5*, *Cldn5*, *Pecam1*, *Cdkn2a*, *Cdkn1a*, *Cdkn1b*, *Ccnd1*, *Cdk2*, *Cdk4*, *Tgfb1*, *Tgfb2*, *Tgfb3*, *Bmp2*, *Bmp4*, *Nog*, *Chrd*, *Cer1*, *Fgf1*, *Fgf8*, *Wnt1*, *Wnt3a*, *Wnt5a*, *Wnt10b*, *Dkk1* and *Pgf* were used along with TaqMan Gene Expression Master Mix (Applied Biosystems) to perform qPCR. Gene expression assays were normalized to the endogenous VIC-conjugated *Actb* probes as control. For the analysis of *Bgalp*, *Ibsp* and *Vegfa* expression, RNA was isolated from the whole bone samples of mutants and littermate control animals. Freshly dissected femurs were immediately flushed (to remove haematopoietic cells), crushed and then lysed in lysis buffer. RNA isolation, cDNA preparation and qPCR were performed as described above.

Flow cytometry. For flow cytometric analysis of endothelial cells, femurs and tibiae were collected, cleaned thoroughly to remove the adherent muscles and then crushed in ice-cold PBS with a mortar and pestle. Whole bone marrow was digested with collagenase at 37 °C for 20 min. Cells were counted and equal numbers of cells were subjected to immunostaining. Cells were stained with biotin-coupled CD45 (BD, 553077) and Ter119 (BD, 559971) antibodies for 45 min. After washing in PBS, cells were stained with Streptavidin PE-Cy5 (BD, 554062) and CD31-conjugated with Alexa Fluor 488 (R&Dsystems, FAB3628G) for 45 min. After washing, data was acquired on FACS Aria II or FACS Canto flow cytometers and analysed using FACSDiva software (Version 6.0, BD Bioscience). For analysis and sorting of $CD31^{hi}$ $Emcn^{hi}$ endothelial cells, first standard quadrant gates were set. Subsequently, to differentiate $CD31^{hi}Emcn^{hi}$ cells from the total double positive cells in quadrant 2, gates were arbitrarily set at $>10^4$ log Fl-4 (CD31-APC) fluorescence and $>10^4$ log Fl-2 (Endomucin-PE) fluorescence.

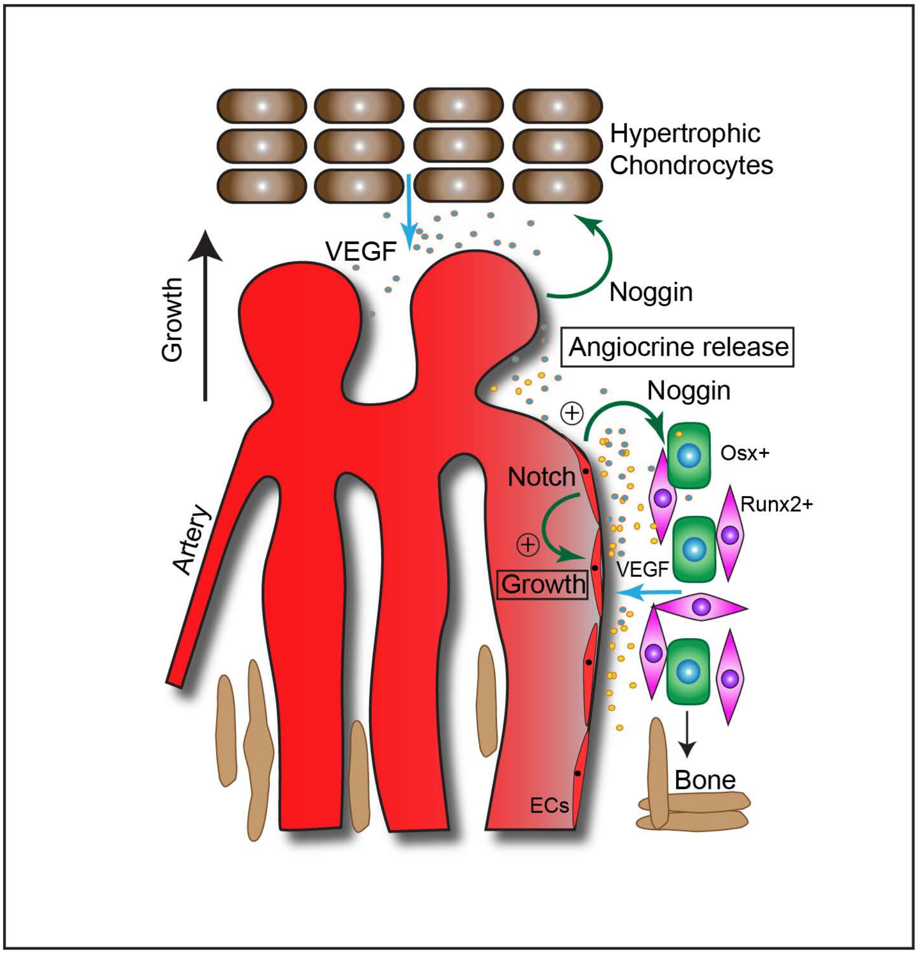
Micro-CT analysis, histomorphometry and electron microscopy. Tibiae were collected from mutants and their littermate controls; the attached soft tissue in the bone was removed thoroughly and fixed in 2% paraformaldehyde. The fixed tibiae were analysed using micro-CT (μCT 35) and software IPL V5.15 at Scanco Medical AG, Switzerland. A voxel size of 12 μm was chosen in all three spatial dimensions. For each sample, 148 from 232 slices were evaluated covering a total of 1.776 mm at a voltage of 70 kVp, intensity 114 μA, and integration time 1200 ms.

Calcein double labelling was performed to calculate bone formation rate (BFR) and mineral apposition rate (MAR). Mice were given intraperitoneal injections of 10 mg per kg calcein (Sigma, C0875) dissolved in 2% sodium bicarbonate solution at the tenth day and third day before euthanasia. Bones were fixed in 4% PFA, embedded in 8% gelatin and 2% PVP and cryosectioned. Single plane images were acquired from the sections using LSM 780 (Carl Zeiss). Sections were stained with von Kossa method to assess mineralized bone. Representative images show cortical bone (diaphysis about 3 mm proximally from the growth plate). Mineral apposition rates were calculated from both cortical and trabecular bones. The latter were measured at the proximal part of trabeculae relative to the angiogenic front in the metaphysis. Osteoclast surface/bone surface (Oc. S/BS; %) and osteoclast number/bone perimeter (No. Oc./B. Pm) were calculated based on Calcitonin receptor staining of bone sections.

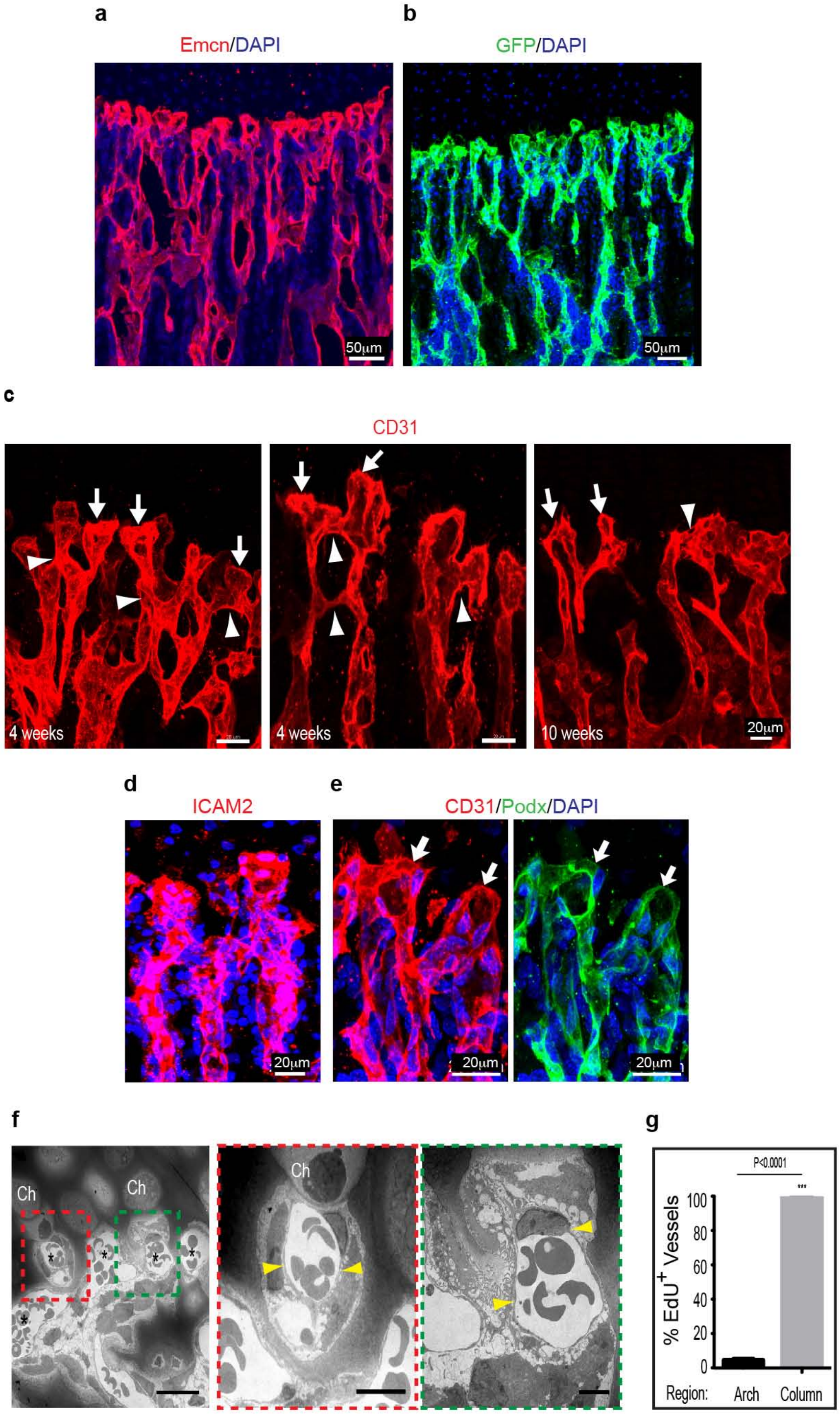
For electron microscopy, freshly collected femurs were processed similar to the protocol published earlier[29]. In brief, fresh bone was directly transferred into 4% paraformaldehyde, 0.5% glutaraldehyde in 0.1 M cacodylate buffer (pH 7.2), cut and fixed initially for 2 h at room temperature and overnight at 4 °C. Bone was decalcified in 5% EDTA in 0.1 M cacodylate buffer (pH 7.2) for 3 days. The sample was post-fixed for 1.5 h in 1% $OsmO_4$, 1.5% potassium ferrocyanide. Dehydration and embedding in epon was done under a slight vacuum. In all, 70 nm ultrathin sections of the sample were taken in the area of interest and collected on filmed single slot copper grids (Leica UC6 ultramicrotome, Vienna, Austria). Sections got counterstained with uranyl acetate and lead, and were analysed at 80 kV on a FEI-Tecnai 12 electron microscope (FEI, Eindhoven, The Netherlands). Pictures were taken with imaging plates (Ditabis, Pforzheim, Germany).

Statistics. All data are presented as mean ± s.e.m. The significance of difference in the mean values was determined using two-tailed Student's *t* test unless otherwise mentioned. $P < 0.05$ was considered significant. In others, non-parametric one-way ANOVA was performed along with Bonferroni's multiple comparison post-hoc tests to assess statistical significance with a 95% confidence interval. All calculations were performed using GraphPad Prism software. No randomization or blinding was used and no animals were excluded from analysis. Sample sizes were selected on the basis of previous experiments. Several independent experiments were performed to guarantee reproducibility of findings.

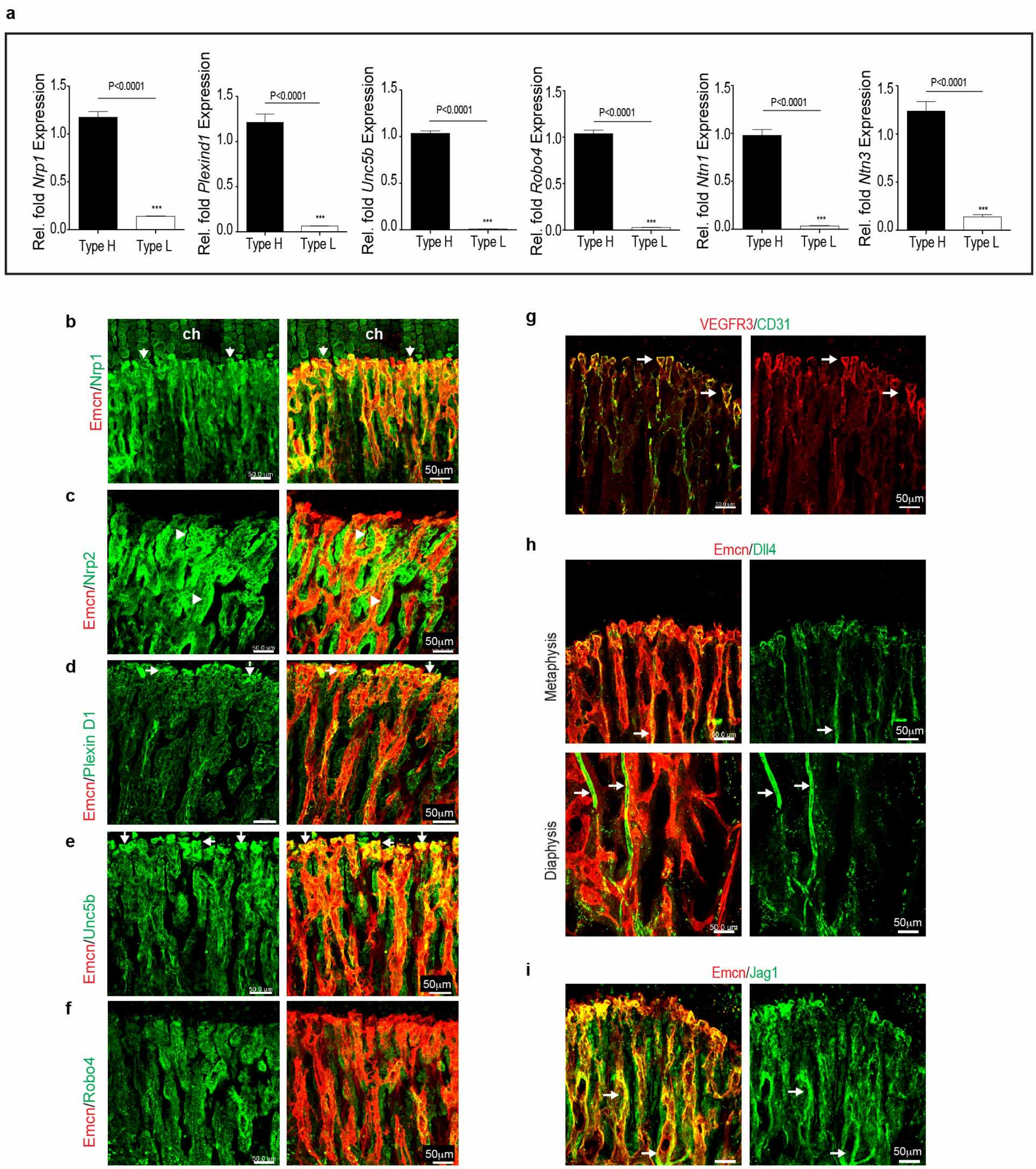
27. Soleimani, M. & Nadri, S. A protocol for isolation and culture of mesenchymal stem cells from mouse bone marrow. *Nature Protocols* **4,** 102–106 (2009).
28. Zhu, H. *et al.* A protocol for isolation and culture of mesenchymal stem cells from mouse compact bone. *Nature Protocols* **5,** 550–560 (2010).
29. Suzuki, H. *et al.* Histological evidence of the altered distribution of osteocytes and bone matrix synthesis in klotho-deficient mice. *Arch. Histol. Cytol.* **68,** 371–381 (2005).



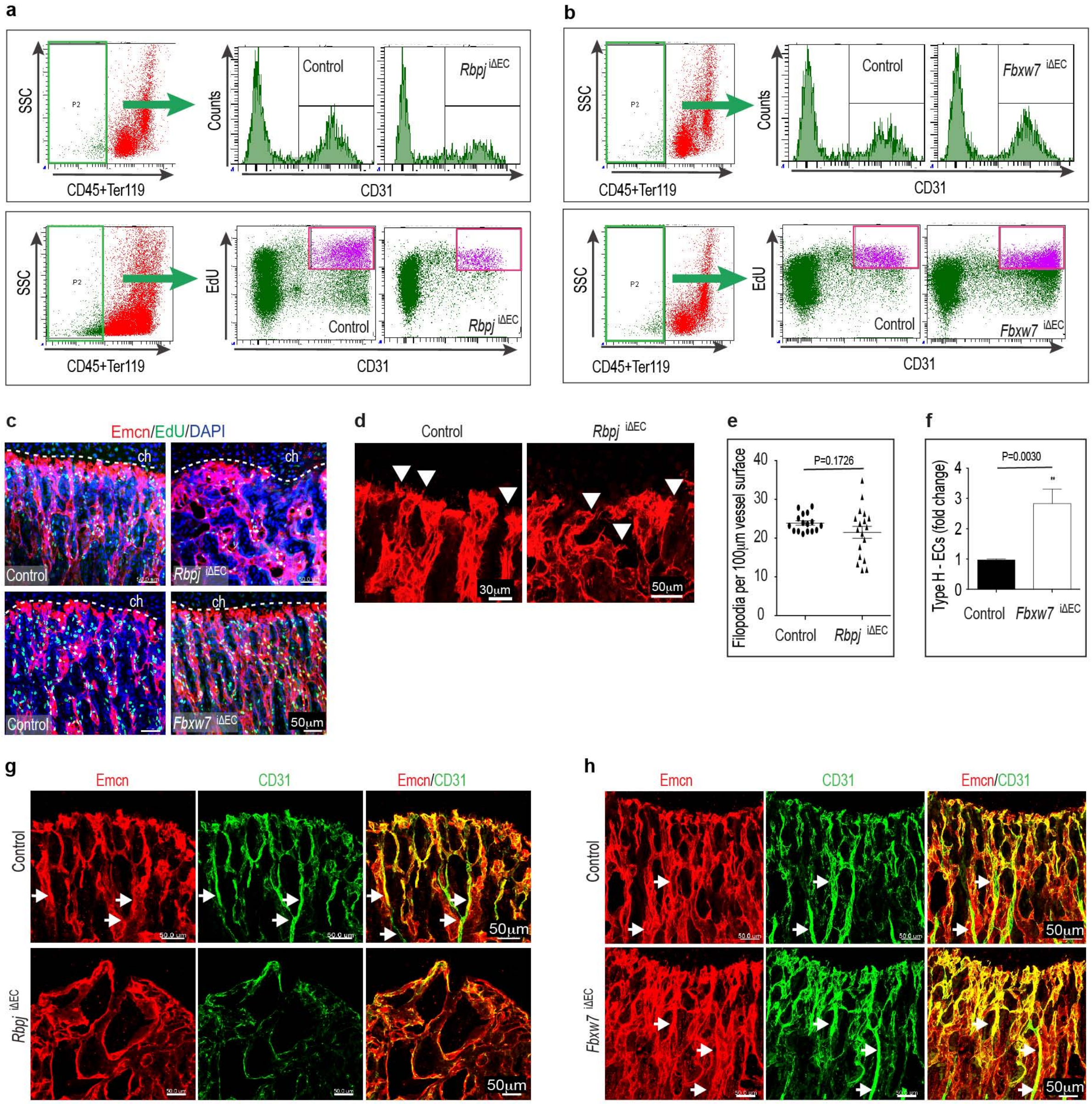
Extended Data Figure 1 | Schematic representation of key findings. Organization and role of growing vessels in the regulation of osteogenesis in postnatal long bone. Endothelial columns, which are embedded between segments of forming trabecular bone in the metaphysis, are interconnected by arches at their distal end. Blind-ended, lumen-containing protrusions extend from arches towards growth plate chondrocytes, a key source of VEGF-A. Endothelial Notch signalling promotes endothelial cell proliferation and vessel growth in bone, which is the opposite of its role in other tissues. Notch activity in endothelial cells is also required for endothelial Noggin expression, controls the differentiation of perivascular osteoprogenitor cells and thereby osteogenesis. Endothelial Notch signalling and Noggin also promote chondrocyte maturation and hypertrophy, which affects angiogenesis through VEGF-A expression. These signalling interactions between different cell types couple angiogenesis and osteogenesis.



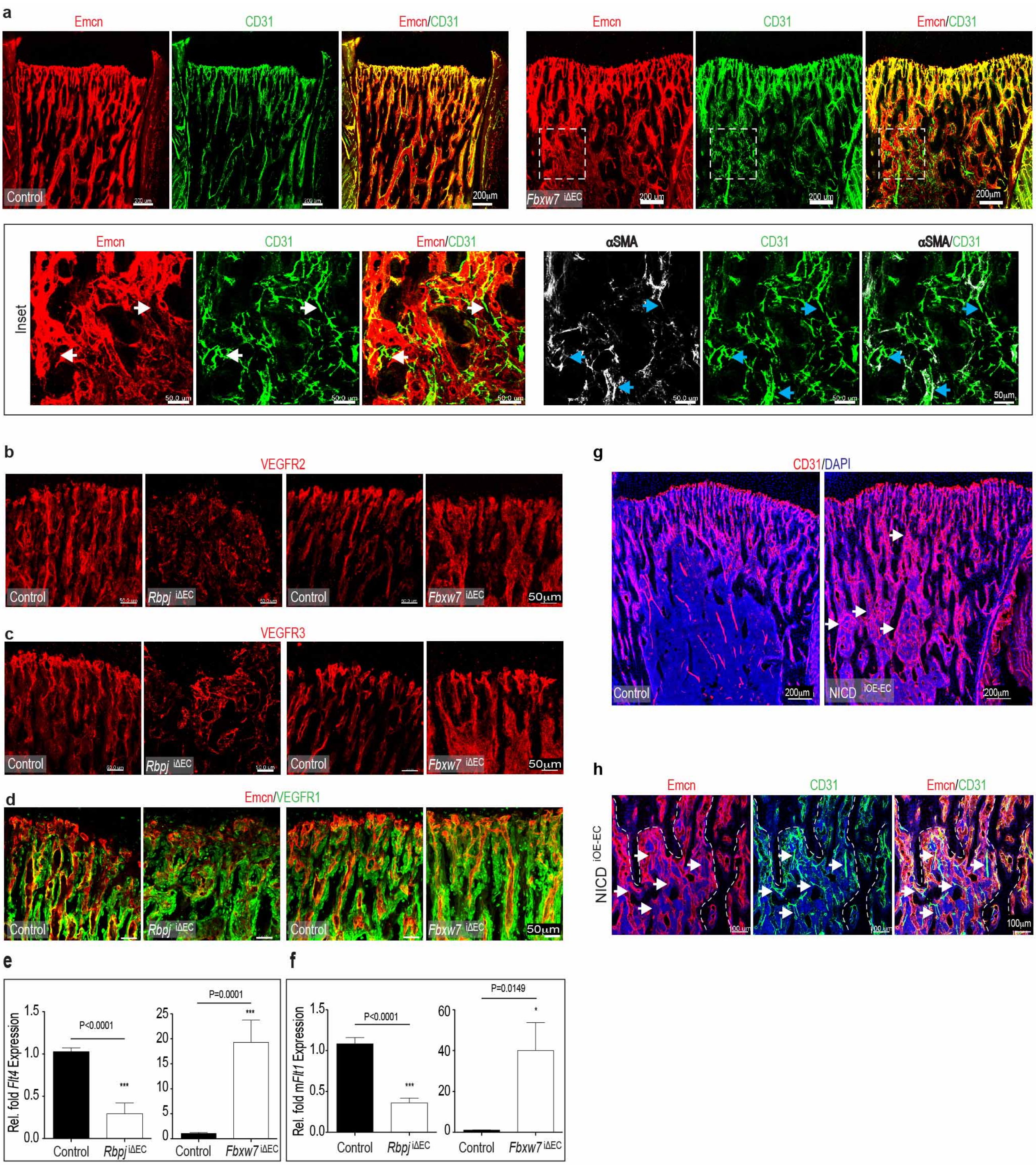
Extended Data Figure 2 | Vessel growth in the postnatal metaphysis. **a**, **b**, Organisation of distal vessels in the metaphysis of 4-week-old tibia. Endothelial cells were visualized by anti-Endomucin (Emcn, red) immunostaining (**a**) or GFP expression (green) in *Cdh5(PAC)-CreERT2 Rosa26-mT/mG* double transgenic mice (**b**) at 4 weeks of age. Note blunt appearance of most distal vessels (top) in proximity of growth plate chondrocytes. Nuclei, blue (DAPI). **c**, Maximum intensity projection showing the organization of distal, CD31-immunostained (type H) vessels in the tibial metaphysis at the indicated ages. Note emergence of blunt and blind-ended bulb-like protrusions (arrows) from arch vessels (arrowheads) at the distal end of endothelial columns. **d**, Confocal image showing expression of intercellular adhesion molecule 2 (ICAM2, red), a marker of lumenized vessels, in distal vessels of the tibial metaphysis. Nuclei, blue (DAPI). **e**, Podocalyxin (Podx, green), a sialoglycoprotein marking the apical surface of endothelial cells and thereby the vascular lumen (arrows), is present on the most distal, CD31-positive (red) vessel structures in the metaphysis. Nuclei, blue (DAPI). **f**, Transmission electron microscopy confirming the lumenized nature of distal vessels close to growth plate chondrocytes (ch). Yellow arrowheads indicate thin endothelial cells lining vessels. **g**, Quantitation of EdU labelled (proliferating) endothelial cells in the metaphysis of long bone. EdU^+ endothelial cells were predominantly present in columnar vessels and were comparably rare ($\leq$5%) in vascular arches ($n = 5$ mice from 3 independent experiments). Error bars, $\pm$ s.e.m. P values, two-tailed unpaired t-test.



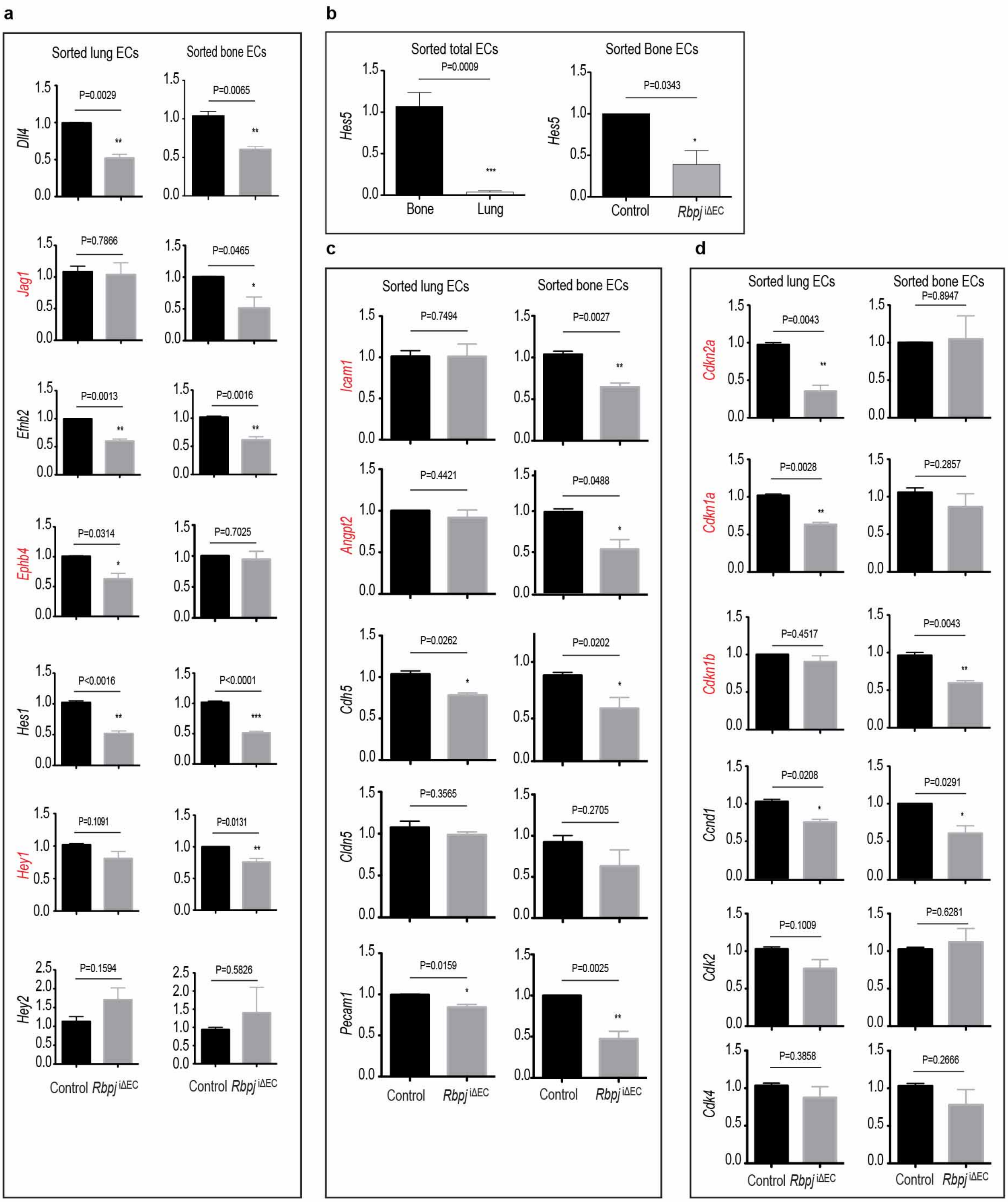
Extended Data Figure 3 | Marker expression in arch and bulb endothelial cells. **a**, Quantitative real time PCR analysis of transcripts encoding vessel guidance molecules in type H and type L endothelial cells isolated by FACS from 4-week-old femur. Note high levels of transcripts for Netrin-1 (*Ntn1*), Netrin-3 (*Ntn3*), Neuropilin 1 (*Nrp1*), Plexin D1 (*Plxnd1*), Unc5b and Robo4 in type H relative to type L endothelium ($n = 5$ mice from 4 independent experiments). Error bars, ± s.e.m. P values, two-tailed unpaired t-test. **b**–**f**, Confocal images showing immunostaining of vessel guidance molecules such as Neuropilin 1 (**b**, Nrp1), Neuropilin 2 (**c**, Nrp2), Plexin D1 (**d**), Unc5b (**e**), Robo4 (**f**) in metaphyseal (type H) vessels and surrounding mesenchymal cells. High levels of Nrp1 were detected in type H endothelial cells (arrows), but comparably lower expression was also observed in vessel-associated osteoprogenitors and growth plate chondrocytes (ch). Nrp2 is highly expressed in perivascular osteoprogenitors (arrowheads). High expression of PlexinD1 and Unc5b, and lower levels of Robo4 were detected in type H vessel bulbs and columns (arrows). **g**, Maximum intensity projection showing high expression of VEGFR3 in endothelial bulbs and arches (white arrows) compared to columns. VEGFR3 was absent in the arteries. **h**, Image showing Dll4 expression in type H endothelial cells forming bulbs and arches in 4-week-old tibial metaphysis. With the exception of arteries (arrows), Dll4 expression was low in vessels of the diaphysis. **i**, Jag1 expression was detected in endothelial cells (white arrows) and surrounding mesenchymal cells.



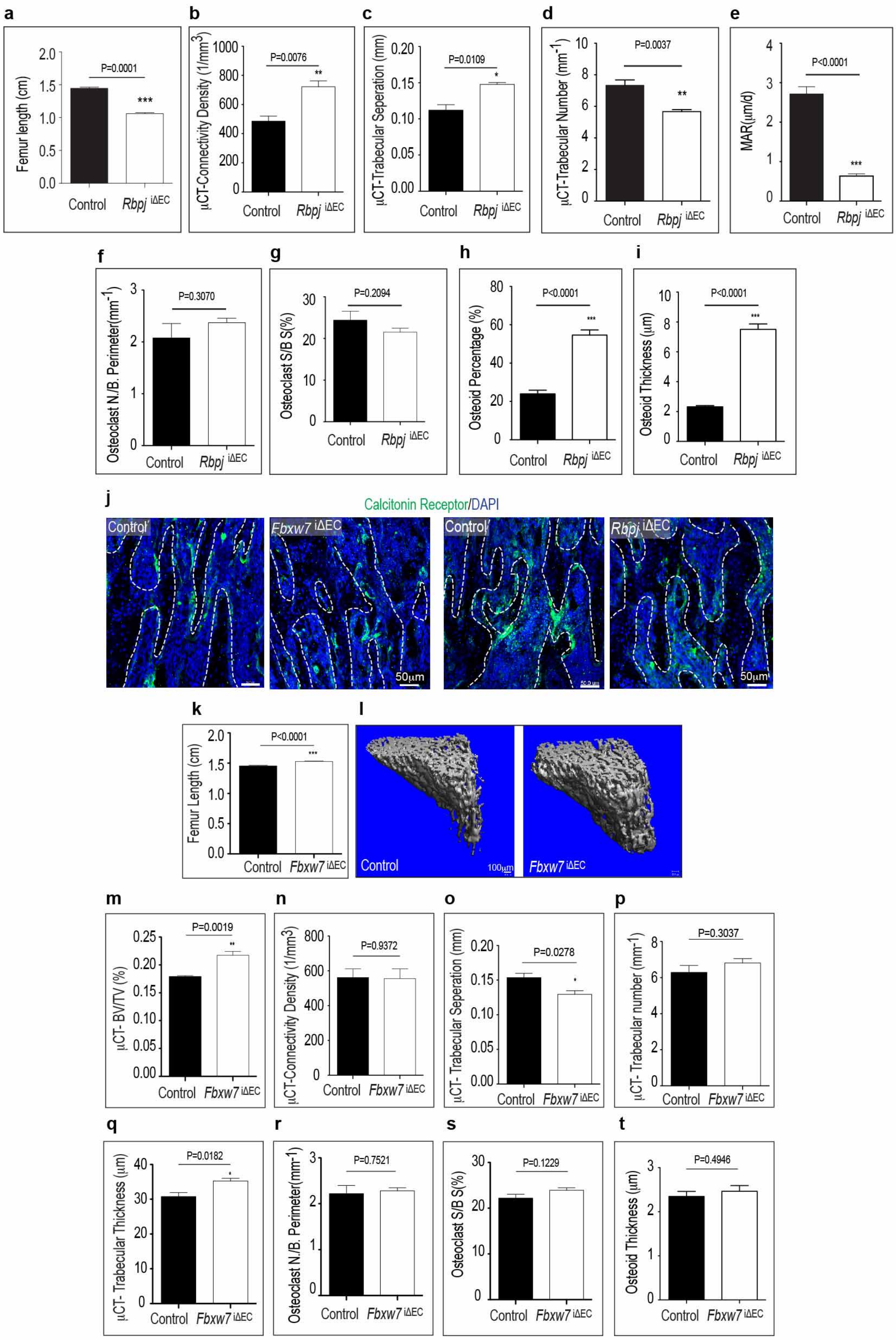
Extended Data Figure 4 | EC numbers, proliferation and type H vessels in Notch mutants. **a**, **b**, Representative flow-cytometric graph plots showing the quantification of total and EdU-labelled endothelial cells from *Rbpj*$^{i\Delta EC}$ (**a**) or *Fbxw7*$^{i\Delta EC}$ (**b**) mutant bones or corresponding littermate controls, as indicated. CD31^{+} CD45^{-} Ter119^{-} endothelial cells were substantially reduced in *Rbpj*$^{i\Delta EC}$ mutants (**a**). Conversely, endothelial cell number and proliferation were increased after inactivation of *Fbxw7* (**b**). **c**, Maximum intensity projections of EdU-labelled (green fluorescence) tibial metaphysis showing proliferating cells in 4-week-old Notch gain-of-function (*Fbxw7*$^{i\Delta EC}$) and loss-of-function (*Rbpj*$^{i\Delta EC}$) mutants and corresponding littermate controls. Endothelial cells were visualized by Emcn immunostaining (red) and nuclei by DAPI (blue). **d**, Confocal images of Emcn (red) immunostained endothelial distal columns and arches next to growth plate chondrocytes (ch) highlighting the distribution of filopodia (arrowheads). Note that filopodia were directed towards chondrocytes in controls whereas directionality was compromised in *Rbpj*$^{i\Delta EC}$ mutants. **e**, Quantitation of filopodia indicating higher variability in *Rbpj*$^{i\Delta EC}$ samples ($n = 6$ mice from 4 independent litters). Error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-test. **f**, Quantitative analysis of type H vessels in 4-week-old *Fbxw7*$^{i\Delta EC}$ mutants and littermate controls ($n = 6$ mice from 4 independent litters). Error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-test. **g**, **h**, Maximum intensity projections of CD31 (green) and Endomucin (red) immunostained vessels in the metaphysis of 4-week-old *Rbpj*$^{i\Delta EC}$ (**g**) and *Fbxw7*$^{i\Delta EC}$ mutants (**h**). Note loss of type H vessels and decreased CD31 (green) staining in the *Rbpj*$^{i\Delta EC}$ metaphysis, while CD31^{+} vessels were extended after endothelial-cell-specific inactivation of *Fbxw7*. Arrows indicate arteries.



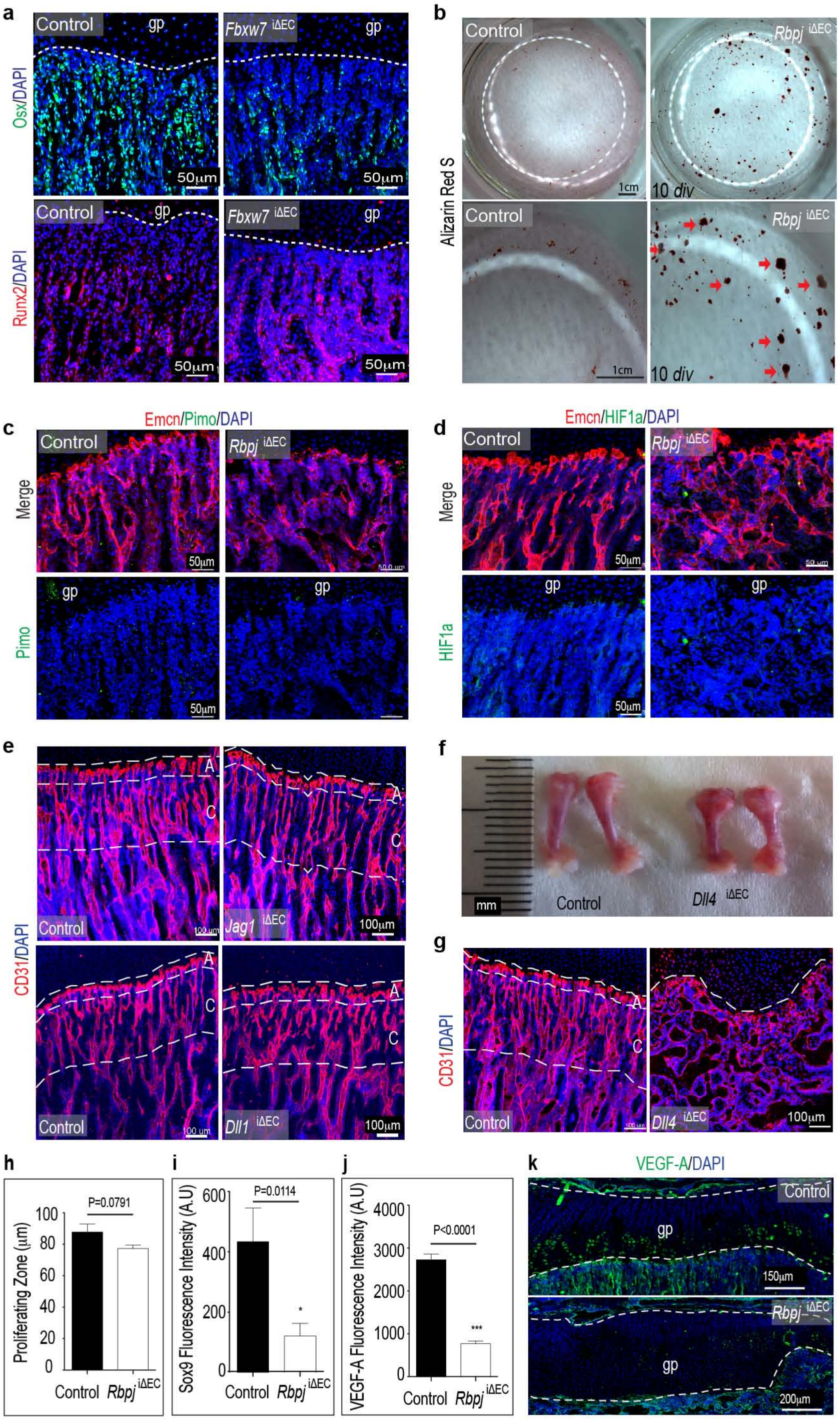
Extended Data Figure 5 | Arterial specification and VEGF receptor expression in Notch mutants. **a**, Maximum intensity projections of Emcn (red) and CD31 (green) immunostained sections of 4-week-old tibia showing increase in CD31^{+} (type H) vessels in $Fbxw7^{i\Delta EC}$ mice. Mutants also displayed numerous small CD31^{+} Endomucin^{-} arterioles (white arrows), which were associated with αSMA^{+} cells (blue arrows in bottom panels). **b**–**d**, Confocal images showing 4-week-old $Rbpj^{i\Delta EC}$ or $Fbxw7^{i\Delta EC}$ mutant, or corresponding littermate control metaphysis after immunostaining for different VEGF receptors. VEGFR2 was highest on control arches, bulbs and arteries and staining was strongly reduced in the $Rbpj^{i\Delta EC}$ metaphysis (**b**). VEGFR3 immunostaining decorated arches and bulb protrusion but was absent in the arteries. Staining was reduced in $Rbpj^{i\Delta EC}$ mutants but enhanced in $Fbxw7^{i\Delta EC}$ vessels (**c**). Expression of VEGFR1 was predominantly found on perivascular mesenchymal and osteoprogenitor cells. Expression in these populations was not appreciably altered (**d**). **e**, **f**, qPCR analysis of sorted endothelial cells from Notch loss-of-function ($Rbpj^{i\Delta EC}$, **e**) and gain-of-function ($Fbxw7^{i\Delta EC}$, **f**) mice. In endothelial cells, Notch positively regulated transcripts for the receptors VEGFR2 (*Kdr*), VEGFR3 (*Flt4*) and membrane-anchored VEGFR1 (*mFlt1*). In contrast, expression of soluble Flt1 (*sFlt1*), a known antagonist of VEGF signalling, was increased in $Rbpj^{i\Delta EC}$ endothelial cells and significantly reduced in $Fbxw7^{i\Delta EC}$ cells ($n = 4$ mice from 4 independent litters). Error bars, ± s.e.m. P values, two-tailed unpaired t-test. **g**, Increased formation of CD31^{+} (red) vessels in the metaphysis of 4-week-old Notch gain-of-function mice after endothelial-cell-specific overexpression of active Notch ($NICD^{iOE\text{-}EC}$). Nuclei, blue (DAPI). **h**, Confocal images showing extensive formation of CD31^{+} (green) Emcn^{-} (red) arterioles (arrows) in the $NICD^{iOE\text{-}EC}$ metaphysis. Nuclei, DAPI (blue).



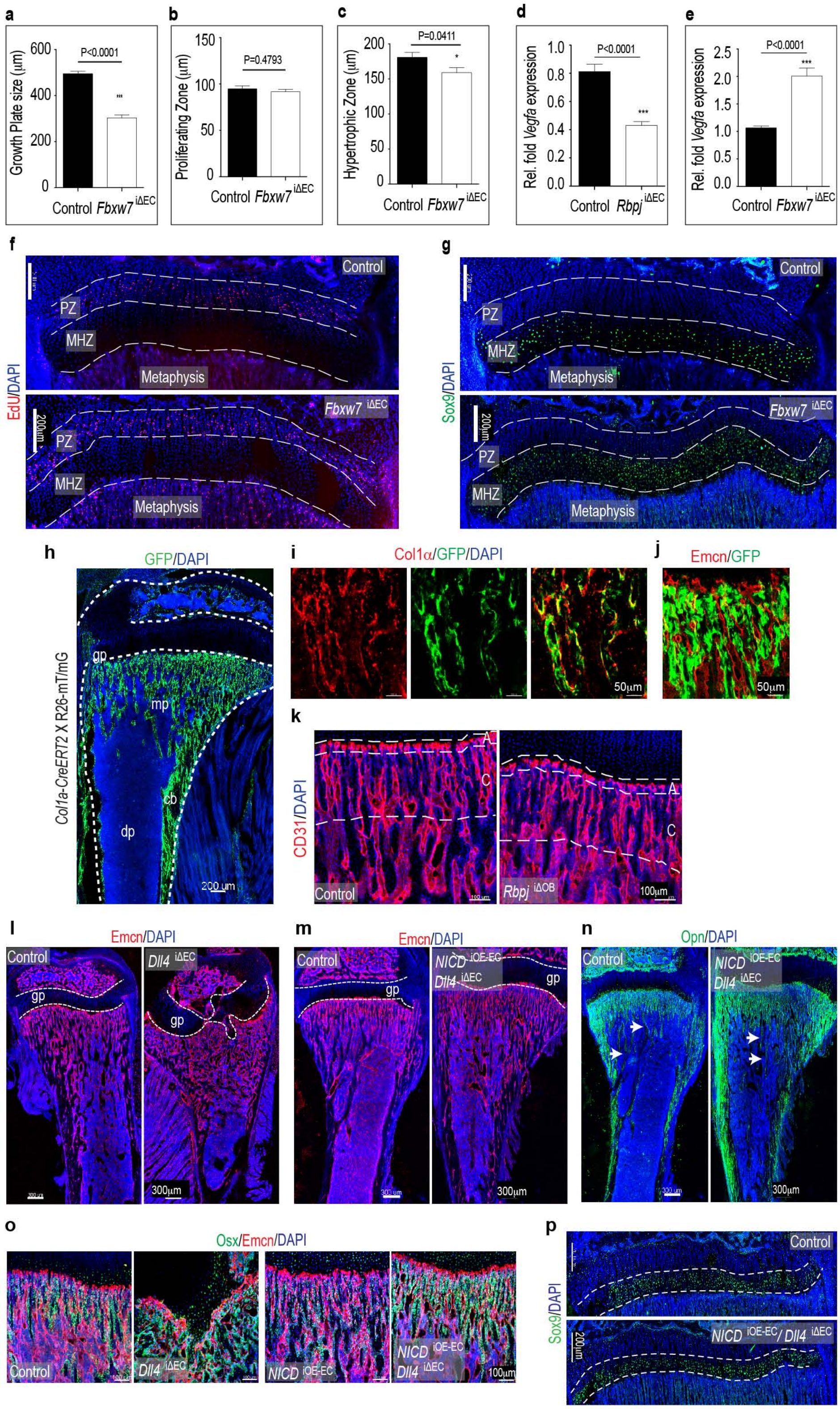
Extended Data Figure 6 | Gene expression in *Rbpj* mutant bone and lung endothelial cells. **a**, qPCR analysis of freshly isolated endothelial cells from control and *Rbpj*$^{i\Delta EC}$ mutant lung or bone. Relative fold mRNA expression of Notch target genes such as *Dll4*, *Jag1*, *Efnb2*, *EphB4*, *Hes1*, *Hey1*, *Hey2* is shown. Differently expressed genes are marked in red ($n = 4$ mice from 4 independent litters). Error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-test. **b**, Quantitative PCR analysis of expression the Notch target gene *Hes5* in sorted endothelial cells from 4-week-old wild-type bone and lung (left), and from *Rbpj*$^{i\Delta EC}$ mutant and littermate control bone samples (right). Note very low expression of *Hes5* in lung compared to bone endothelial cells and its reduction in the *Rbpj*$^{i\Delta EC}$ mutant bone endothelium. Data represent relative fold mRNA expression ($n = 4$ mice from 4 independent litters). Error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-test. **c**, qPCR results for transcripts encoding intercellular adhesion molecule 1 (*Icam1*), angiopoietin 2 (*Angpt2*), VE-cadherin (*Cdh5*), and CD31 (*Pecam1*). Data represent relative fold mRNA expression. Differently expressed genes are marked in red ($n = 4$ mice from 4 independent litters). Error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-test. **d**, Relative fold mRNA expression of cyclin-dependent kinase inhibitor 2A (*Cdkn2a*), cyclin-dependent kinase inhibitor 1A (*Cdkn1a*), cyclin-dependent kinase inhibitor 1B (*Cdkn1b*), cyclin D1 (*Ccnd1*), cyclin-dependent kinase 2 (*Cdk2*), and cyclin-dependent kinase 4 (*Cdk4*) in control or *Rbpj*$^{i\Delta EC}$ mutant lung and bone endothelial cells ($n = 4$ mice from 4 independent litters). Error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-test.



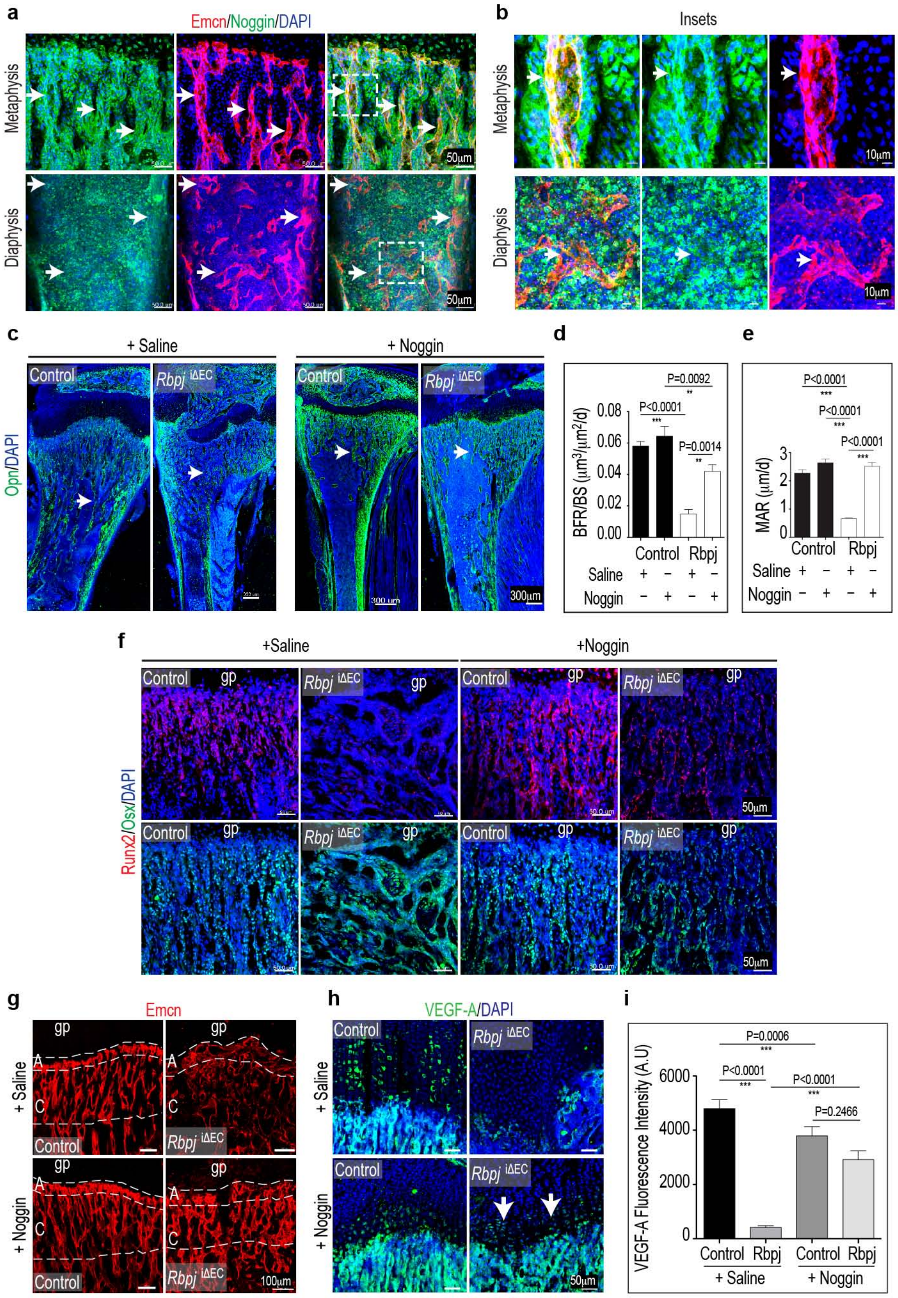
Extended Data Figure 7 | Endothelial Notch controls osteogenesis. **a–i**, Analysis of bone parameters in 4-week-old *Rbpj*$^{i\Delta EC}$ mutants and control littermates. Data on femur length are shown in panel **a**. Connectivity density (number of connections per unit volume, **b**), trabecular bone separation (size of space separating trabeculae, **c**), trabecular number (number of trabeculae per mm length, **d**) were obtained by µ-CT. Mineral apposition rates (MAR, **e**) were calculated using calcein double labelling. Data on osteoclast number per bone perimeter (**f**), osteoclast surface per bone surface (**g**), osteoid percentage (**h**) and osteiod thickness (**i**) are based on histomorphometrical characterization of control and mutant samples ($n = 6$ mice from 4 independent litters). Error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-test. **j**, Osteoclasts were identified by immunostaining for Calcitonin receptor (green). Maximum intensity projection of Calcitonin receptor stained metaphysis region of tibial sections of Notch gain-of-function (*Fbxw7*$^{i\Delta EC}$) and loss-of-function (*Rbpj*$^{i\Delta EC}$) mutants along with their respective littermate controls. Nuclei, DAPI (blue). **k–t**, Analysis of bone parameters in 4-week-old *Fbxw7*$^{i\Delta EC}$ mutants and control littermates. Femur length is shown in **k**. µ-CT data (**l**) were used for the analysis of bone density (bone volume/total volume; BV/TV, **m**), connectivity density (number of connections per unit volume, **n**), trabecular bone separation (size of space separating trabeculae, **o**), trabecular number (number of trabeculae per mm length, **p**), and trabecular thickness (**q**). Data on osteoclast number per bone perimeter (**r**), osteoclast surface/bone surface (Osteoclast S/B S) (**s**), and osteiod thickness (**t**) are based on histomorphometrical characterization of control and mutant samples ($n = 6$ mice from 4 independent litters). Error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-test.



Extended Data Figure 8 | Endothelial Notch controls mesenchymal cell differentiation. **a**, Confocal images showing immunostaining for mesenchymal lineage markers Osx (green) and Runx2 (red) in *Fbxw7*$^{i\Delta EC}$ and littermate control tibiae, as indicated. Nuclei, DAPI (blue). Note reduction of Osx^{+} cells and increased Runx2+ population after targeting of *Fbxw7* in endothelial cells. **b**, Analysis of differentiation capacity of primary mesenchymal cells isolated from *Rbpj*$^{i\Delta EC}$ or control bone. Alizarin Red S staining after 10 days *in vitro* (10 div) showed that mesenchymal cells isolated from endothelial-cell-specific *Rbpj*$^{i\Delta EC}$ mutants generated mineral nodules (red arrows) prematurely in comparison to controls. **c, d**, Maximum intensity projection of stained *Rbpj*$^{i\Delta EC}$ and control tibiae. Labelling of hypoxic cells with pimonidazole (Pimo, green) showed no appreciable differences between control and *Rbpj*$^{i\Delta EC}$ samples (**c**). Likewise, anti-HIF1α immunostaining (green) was comparable in the control and *Rbpj*$^{i\Delta EC}$ metaphysis (**d**). Endothelial cells (Emcn antibody staining, red); nuclei, DAPI (blue). **e**, Representative confocal images showing that endothelial-cell-specific gene targeting of Jagged1 (Jag1) or Delta-like 1 (Dll1) did not lead to appreciable alterations in *Jag1*$^{i\Delta EC}$ or *Dll1*$^{i\Delta EC}$ mutant metaphyseal vasculature at 4 weeks. Dashed lines indicate vessel columns (C) and distal arches (A). Endothelial cells, CD31 (red); nuclei, DAPI (blue). **f**, Smaller size of freshly isolated femurs from 4 week-old *Dll4*$^{i\Delta EC}$ mutants relative to littermate controls. **g**, Disrupted organization of vessel arches (A) and columns (C) in the *Dll4*$^{i\Delta EC}$ metaphyseal vasculature. Endothelial cells, CD31 (red); nuclei, DAPI (blue). **h**, Quantitative analysis of size of chondrocyte proliferating zone in littermate control and *Rbpj*$^{i\Delta EC}$ mutants ($n = 6$ mice from 4 independent litters). Error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-test. **i, j**, Quantification of Sox9 (**i**) and VEGF-A (**j**) fluorescence intensity in immunostained sections of *Rbpj*$^{i\Delta EC}$ mutants showing decreased signals in mutant growth plates ($n = 4$ mice from 3 independent litters). Error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-test. **k**, Maximum intensity projection of stained *Rbpj*$^{i\Delta EC}$ and littermate control tibiae showing decreased VEGF-A immunosignals.


a
Growth Plate size (μm)
P<0.0001
Control Fbxw7 iΔEC
b
Proliferating Zone (μm)
P=0.4793
Control Fbxw7 iΔEC
c
Hypertrophic Zone (μm)
P=0.0411
Control Fbxw7 iΔEC
d
Rel. fold Vegfa expression
P<0.0001
Control Rbpj iΔEC
e
Rel. fold Vegfa expression
P<0.0001
Control Fbxw7 iΔEC
f
EdU/DAPI
Control
PZ
MHZ
Metaphysis
Fbxw7 iΔEC
g
Sox9/DAPI
h
GFP/DAPI
Col1a-CreERT2 X R26-mT/mG
gp
mp
cb
dp
200 μm
i
Col1α/GFP/DAPI
50μm
j
Emcn/GFP
k
CD31/DAPI
A
C
Rbpj iΔOB
100μm
l
Emcn/DAPI
Dll4 iΔEC
300μm
m
NICD iOE-EC
Dll4 iΔEC
n
Opn/DAPI
o
Osx/Emcn/DAPI
NICD iOE-EC
p
NICD iOE-EC/ Dll4 iΔEC

Extended Data Figure 9 | Endothelial-cell-dependent regulation of growth plate and bone. **a**–**c**, Quantitative analysis of *Fbxw7*$^{i\Delta EC}$ mutant growth plates (**a**), proliferating zones (**b**) and maturation/hypertrophy zones (**c**) relative to control littermates ($n = 4$ mice from 3 independent litters). Error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-test. **d**, **e**, Quantitative PCR analysis of *Vegfa* transcripts level in the bones (without marrow cells) of Notch mutants (*Rbpj*$^{i\Delta EC}$ and *Fbxw7*$^{i\Delta EC}$) and respective littermate controls ($n = 4$ mice from 3 independent litters). Error bars, ± s.e.m. *P* values, two-tailed unpaired *t*-test. **f, g**, Maximum intensity projection of stained *Fbxw7*$^{i\Delta EC}$ and littermate control tibiae. Note slight reduction of mutant growth plate and maturation/hypertrophy zone (MHZ). EdU (red) labelling (**f**) marks mitotic chondrocytes in the proliferating zone (PZ). Sox9 (green) immunosignals (**g**) label maturing and hypertrophic chondrocytes. Dashed lines mark borders of PZ and MHZ. Nuclei, DAPI (blue). **h**, Maximum intensity projection of tile scanned tibial section of *Col1a1-CreERT2 ROSA26-mT/mG* double transgenic mice. GFP expression (green) indicates Cre-mediated recombination. **i**, Maximum intensity projections showing overlapping GFP (green) and Col1α immunostaining (red). **j**, Maximum intensity projections showing osteoblasts expressing GFP (green). GFP signals were not seen in Emcn immunostained endothelial cells (red). **k**, Normal arrangement of CD31-stained endothelial columns (C) and arches (A) after *Col1a1-CreERT2*-mediated inactivation of *Rbpj* in cells of the osteoblast lineage (*Rbpj*$^{i\Delta OB}$). Nuclei, DAPI (blue). **l**, **m**, Maximum intensity projections of tile-scanned tibia sections after anti-Emcn (red) immunostaining. Nuclei, DAPI (blue). Note profound disorganization of the *Dll4*$^{i\Delta EC}$ metaphyseal vasculature (**l**) and restoration by simultaneous endothelial-cell-specific overexpression of active Notch in *Dll4*$^{i\Delta EC}$/NICD$^{iOE\text{-}EC}$ double mutant mice (**m**). **n**, Tile scan confocal images of osteopontin (Opn, green) stained control and *Dll4*$^{i\Delta EC}$/NICD$^{iOE\text{-}EC}$ tibiae showing that endothelial NICD expression can rescue trabecular bone defects in the *Dll4*$^{i\Delta EC}$ background (see Fig. 2c for comparison). **o**, Organization of the Emcn-stained metaphyseal vasculature and arrangement of osteoblasts (Osx) in tibia section from 4-week-old control, *Dll4*$^{i\Delta EC}$, NICD$^{iOE\text{-}EC}$, and *Dll4*$^{i\Delta EC}$/NICD$^{iOE\text{-}EC}$ mice, as indicated. Nuclei, DAPI (blue). **p**, Sox9 immunostaining (green) in the epiphyseal region of 4-week-old tibia showing normalized maturing and hypertrophic chondrocytes zone (MHZ) in the growth plate of *Dll4*$^{i\Delta EC}$/NICD$^{iOE\text{-}EC}$ mice. Dashed lines mark borders of MHZ. Nuclei, DAPI (blue).



Extended Data Figure 10 | Notch and angiocrine Noggin production. **a**, Noggin (green) in the wild-type tibial metaphysis was detected in Emcn^{+} (red) endothelial cells as well as surrounding mesenchymal cells. In contrast, the diaphysis contained Noggin^{+} haematopoietic cells, whereas only weak staining was seen in sinusoidal blood vessels. **b**, Panels show higher magnifications of insets in **a**. Nuclei, DAPI (blue). **c**, Confocal tile scans of osteopontin-immunostained tibia sections showing partial restoration of trabecular bone formation in 4-week-old *Rbpj*$^{i\Delta EC}$ mice after administration of recombinant Noggin. Left panels show saline-treated *Rbpj*$^{i\Delta EC}$ mutants and littermate controls. Nuclei, DAPI (blue). **d**, **e**, Noggin treatment restored the bone formation rate (BFR, **d**) and mineral apposition rate (MAR, **e**) in *Rbpj*$^{i\Delta EC}$ long bone to control level ($n = 6$ mice from 4 independent litters). Data represent mean ± s.e.m. One-way ANOVA was performed along with Bonferroni's multiple comparison post-hoc test. **f**, Systemic administration of recombinant Noggin protein reduced the number of Osx^{+} cells (green) and increased Runx2^{+} early osteoprogenitors in the *Rbpj*$^{i\Delta EC}$ metaphysis in comparison to vehicle-treated (saline) mutants. **g**, Maximum intensity projection of Emcn-immunostained control and *Rbpj*$^{i\Delta EC}$ tibia sections after treatment with saline or recombinant Noggin, as indicated. Emcn staining intensity was increased in Noggin-treated *Rbpj*$^{i\Delta EC}$ samples, and the organization of endothelial column and arch structures was partially restored. Dashed lines indicate position of boundaries between endothelial arches (A) and columns (C) as seen in littermate control samples. **h**, **i**, Confocal images of VEGF-A (green) immunostained tibia sections showing growth plate chondrocytes in *Rbpj*$^{i\Delta EC}$ mice after Noggin treatment. Note partial restoration (arrows) of VEGF-A expression in Noggin- but not vehicle control-treated (Saline) *Rbpj*$^{i\Delta EC}$ mutants (**h**). Nuclei, DAPI (blue). Quantification data showing fluorescence intensity (in arbitrary units) of VEGF-A expression recovered in the tibial sections of these animals (**i**) ($n = 6$ mice from 4 independent litters). Data represent mean ± s.e.m. One-way ANOVA was performed along with Bonferroni's multiple comparison post-hoc test.

LETTER

doi:10.1038/nature12974

Two independent transcription initiation codes overlap on vertebrate core promoters

Vanja Haberle[1,2]*, Nan Li[3]*, Yavor Hadzhiev[3], Charles Plessy[4,5], Christopher Previti[6]†, Chirag Nepal[6]†, Jochen Gehrig[3]†, Xianjun Dong[6]†, Altuna Akalin[6]†, Ana Maria Suzuki[4,5], Wilfred F. J. van IJcken[7], Olivier Armant[8], Marco Ferg[8], Uwe Strähle[8], Piero Carninci[4,5], Ferenc Müller[3] & Boris Lenhard[2,9]

A core promoter is a stretch of DNA surrounding the transcription start site (TSS) that integrates regulatory inputs[1] and recruits general transcription factors to initiate transcription[2]. The nature and causative relationship of the DNA sequence and chromatin signals that govern the selection of most TSSs by RNA polymerase II remain unresolved. Maternal to zygotic transition represents the most marked change of the transcriptome repertoire in the vertebrate life cycle[3–6]. Early embryonic development in zebrafish is characterized by a series of transcriptionally silent cell cycles regulated by inherited maternal gene products: zygotic genome activation commences at the tenth cell cycle, marking the mid-blastula transition[7]. This transition provides a unique opportunity to study the rules of TSS selection and the hierarchy of events linking transcription initiation with key chromatin modifications. We analysed TSS usage during zebrafish early embryonic development at high resolution using cap analysis of gene expression[8], and determined the positions of H3K4me3-marked promoter-associated nucleosomes[9]. Here we show that the transition from the maternal to zygotic transcriptome is characterized by a switch between two fundamentally different modes of defining transcription initiation, which drive the dynamic change of TSS usage and promoter shape. A maternal-specific TSS selection, which requires an A/T-rich (W-box) motif, is replaced with a zygotic TSS selection grammar characterized by broader patterns of dinucleotide enrichments, precisely aligned with the first downstream (+1) nucleosome. The developmental dynamics of the H3K4me3-marked nucleosomes reveal their DNA-sequence-associated positioning at promoters before zygotic transcription and subsequent transcription-independent adjustment to the final position downstream of the zygotic TSS. The two TSS-defining grammars coexist, often physically overlapping, in core promoters of constitutively expressed genes to enable their expression in the two regulatory environments. The dissection of overlapping core promoter determinants represents a framework for future studies of promoter structure and function across different regulatory contexts.

Mapping of TSSs using cap analysis of gene expression (CAGE)[8] identified two major promoter classes with respect to TSS precision[10,11]: 'sharp' promoters with one predominant TSS, often associated with a TATA-box that determines TSS selection; and 'broad' promoters with a wider distribution of TSSs, often overlapping with a CpG island. Even with recent reports of the prevalence of known core promoter elements in human promoters[12], the actual mechanism for choosing TSSs within vertebrate promoters in various cell types and conditions remains unknown.

To address developmental-stage-specific promoter usage throughout early embryonic development, we analysed a nucleotide-resolution map of transcription initiation events in the zebrafish genome, generated by CAGE across 12 stages from the unfertilized egg to organogenesis[13] (Fig. 1a). The data revealed numerous cases of promoter dynamics in which maternal messenger RNAs were initiated from different positions than zygotic transcripts, often with shifting of TSS positions within a single promoter (Fig. 1a).

Clustering of individual TSSs by expression profile revealed several major classes of TSS dynamics (Fig. 1b): TSSs present preferentially in maternal (before the mid-blastula transition (MBT)) stages, reflecting maternally inherited transcripts, as opposed to those activated in early or later zygotic stages (post-MBT). Additional clusters included constitutively present TSSs and TSSs with peak activity in the transitional stages, confirming major changes in the zebrafish transcriptome initiated at the MBT[3,14]. An equivalent clustering of entire promoters revealed a similar pattern (Extended Data Fig. 1a). However, promoters with no change in the overall expression level often contained a population of TSSs with very heterogeneous relative usage during development (Fig. 1a and Extended Data Fig. 1b–d).

The observed differential TSS utilization between inherited (maternal) and *de novo* transcribed (zygotic) mRNAs suggested distinct rules for TSS selection acting within the same promoter in the oocyte and the embryo. To reveal the underlying signatures guiding differential promoter interpretation by the maternal and zygotic transcription machinery, we further dissected maternal- and zygotic-specific promoter usage. We first identified a subset of promoters similar to the example in Fig. 1a, showing a significant degree of shifting between maternally and zygotically used TSSs (Extended Data Fig. 2a, b). Our set contained 911 'shifting' promoters whose CAGE signal in pre- and post-MBT stages overlapped by less than 40% (Supplementary Table 1). The TSS shift happened in either direction, but mainly within a narrow window of up to 100 base pairs (bp) (Extended Data Fig. 2c). The preferred maternal and zygotic TSS of those promoters showed antagonistic developmental dynamics, with degradation of inherited maternal transcripts and gradual activation of zygotic ones (Extended Data Fig. 2d).

Aligning sequences of shifting promoters by their maternal dominant TSS revealed a clear enrichment of T- and/or A-containing (WW) dinucleotides ~30 bp upstream of the maternal TSS, hinting at the presence of a functional TATA-like element[15] (Fig. 2a and Extended Data Fig. 2e). By contrast, the zygotic TSS did not show a TATA-like signal in the expected position, but had a sharp SS–WW boundary in local

[1]Department of Biology, University of Bergen, Thormøhlensgate 53A, N-5008 Bergen, Norway. [2]Institute of Clinical Sciences and MRC Clinical Sciences Center, Faculty of Medicine, Imperial College London, Hammersmith Hospital, Du Cane Road, London W12 0NN, UK. [3]School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. [4]RIKEN Omics Science Center, Yokohama, Kanagawa 230-0045, Japan. [5]RIKEN Center for Life Science Technologies, Division of Genomic Technologies, RIKEN Yokohama Campus, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. [6]Computational Biology Unit, Uni Computing, Uni Research AS, University of Bergen, Thormøhlensgate 55, N-5008 Bergen, Norway. [7]Erasmus Medical Center, Center for Biomics, Room Ee679b, Dr Molewaterplein 50, 3015 GE Rotterdam, The Netherlands. [8]Institute of Toxicology and Genetics, Karlsruhe Institute of Technology, Postfach 3640, 76021 Karlsruhe, Germany. [9]Department of Informatics, University of Bergen, Thormøhlensgate 55, N-5008 Bergen, Norway. †Present addresses: German Cancer Research Center (DKFZ), Genomics & Proteomics Core Facility (GPCF), Im Neuenheimer Feld 580/TP3, Heidelberg 69120, Germany (C.Pr.); Broegelmann Research Laboratory, The Gade Institute, University of Bergen, The Laboratory Building, Haukeland University Hospital, N-5021 Bergen, Norway (C.N.); Acquifer AG, Sophienstraße 136, 76135 Karlsruhe, Germany (J.G.); Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA (X.D.); Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland (A.A.).
*These authors contributed equally to this work.

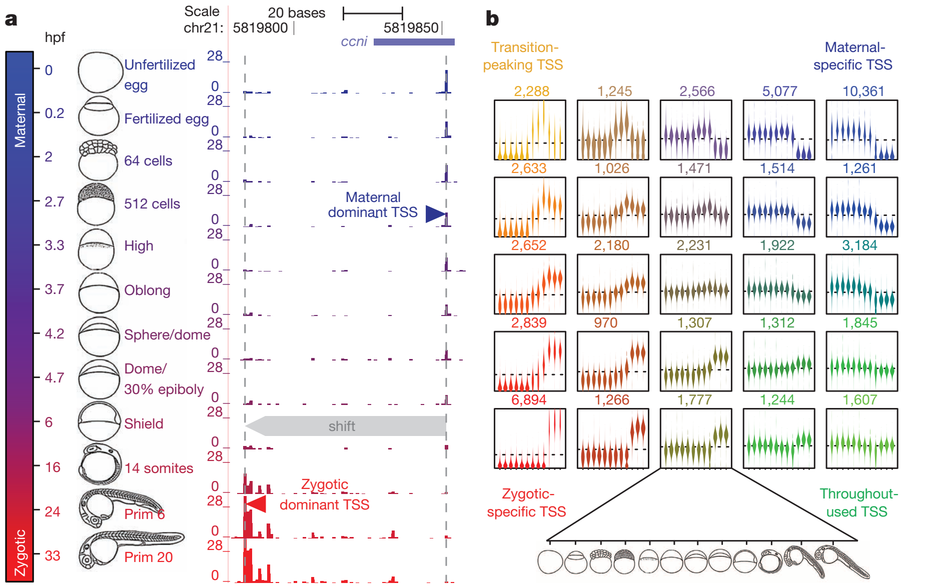


Figure 1 | Dynamics of transcription initiation at 1 bp resolution throughout zebrafish early embryonic development. a, CAGE signal at shifting promoter of cyclin 1 (*ccni*) gene. Colouring from blue to red reflects maternal to zygotic transition. Corresponding zebrafish developmental stages are depicted on the left, with timescale denoting hours post-fertilization (hpf). **b**, Expression profiles obtained by self-organizing map clustering of individual CAGE transcription start sites (CTSSs). Each box represents one cluster, with beanplots showing the distribution of relative expression at different time points for all CTSSs belonging to that cluster (number above the box). Developmental stages, along the *x*-axis for all boxes, are shown at the bottom.

C/G and A/T dinucleotide enrichment precisely aligned ~50 bp downstream of the zygotic TSS (Fig. 2a and Extended Data Fig. 2e). This suggests that there are two fundamentally different sequence signals guiding transcription initiation in the oocyte and the embryo.

Only a small fraction of maternal TSSs (<10%) had a canonical TATA-box motif (Fig. 2b), whereas the majority contained other A/T-rich pentamers (Extended Data Fig. 2f). Motif discovery revealed the presence of an A/T-rich motif (W-box) with lower information content than the canonical TATA-box, but equally positioned 30 bp upstream of the maternal TSS (Fig. 2c). By contrast, the zygotic TSS did not show the presence of a TATA-box or W-box in the expected upstream region (Fig. 2b). This reveals a shift from W-box motif-dependent TSS selection in the maternal transcriptome to zygotic W-box-independent TSS selection; that is, the existence of two major, independent mechanisms for defining transcription initiation acting on the same core promoter.

We conjectured that the uncovered rules for maternal and zygotic TSS selection may apply generally to all constitutively expressed genes, even in the absence of clear TSS shifting. Dinucleotide analysis of all 8,369 constitutively expressed promoters showed the same precise positioning of the W-box signal upstream of the maternal dominant TSS, and alignment of the zygotic TSS with the downstream SS–WW boundary, as seen in the shifting promoters (Fig. 3a and Extended Data Fig. 3a–c). This confirmed a promoterome-wide distinction between determinants that govern TSS selection in the oocyte and the embryo, and demonstrated that complex TSS patterns in constitutively expressed promoters represent readouts of two independent grammars intertwined in the same core promoter regions.

Finally, we showed that exclusively maternal and exclusively zygotic promoters also utilize the corresponding stage-specific TSS selection signals (Extended Data Fig. 3f–h). These results confirm a global change in promoter interpretation that constitutes a central part of the maternal to zygotic transition, with fundamental differences in the TSS selection mechanism used by the transcription machinery in the oocyte and the embryo.

Fixed spacing between the motif and the TSS imposed by W-box-dependent initiation in the oocyte predicts the sharp TSS configuration[10]. The set of maternal broad promoters, which seemingly contradicted the imposed constraints, revealed a novel promoter architecture composed of multiple individual relatively sharp CAGE tag clusters (TCs), each with its associated W-box at a fixed approximately −30 bp position (Extended Data Fig. 4). On the other hand, the exclusively zygotic promoters showed a less constrained distribution of TSSs, revealing the familiar shape of a broad promoter[10], with the majority (>70%)

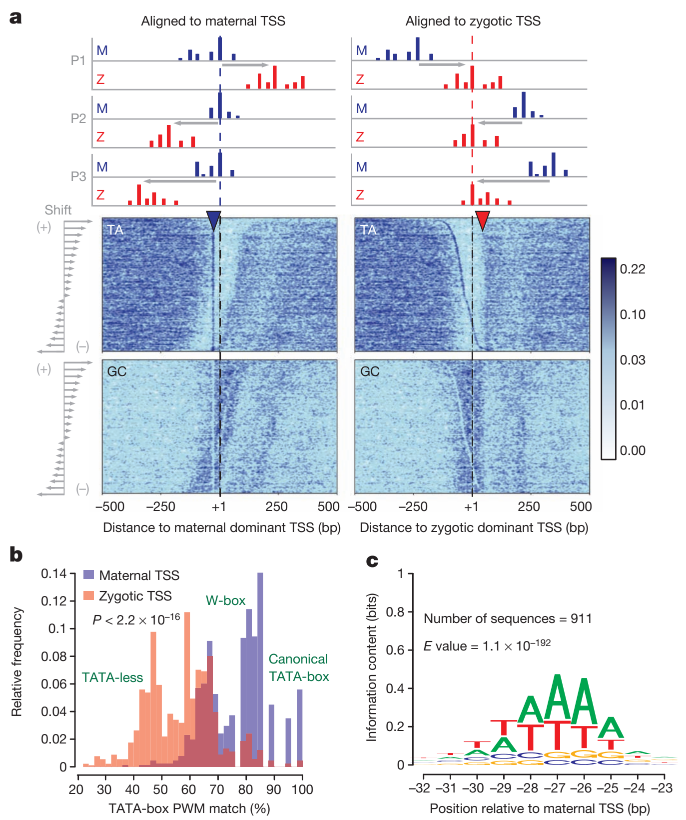


Figure 2 | Sequence signature of a large set of shifting promoters changes markedly during maternal to zygotic transition. a, Dinucleotide density (see Extended Data Fig. 10) at 911 shifting promoters sorted and aligned according to the distance and orientation of the TSS shift (schematics on the top; M, maternal stage; P1, P2, P3, individual promoters; Z, zygotic stage). Promoters were centred at either the maternal (left) or the zygotic (right) dominant TSS. Blue arrowhead indicates TA enrichment at the expected position of the TATA-box; red arrowhead indicates boundary between GC and TA enrichment ~50 bp downstream of zygotic TSS. **b**, Distribution of match (%) to TATA-box in the region −35 to −22 bp upstream of the maternal (blue) and zygotic (red) dominant TSS (*P* value, two-tailed Wilcoxon rank-sum test). PWM, position weight matrix. **c**, Motif obtained by motif discovery upstream of maternal dominant TSS.

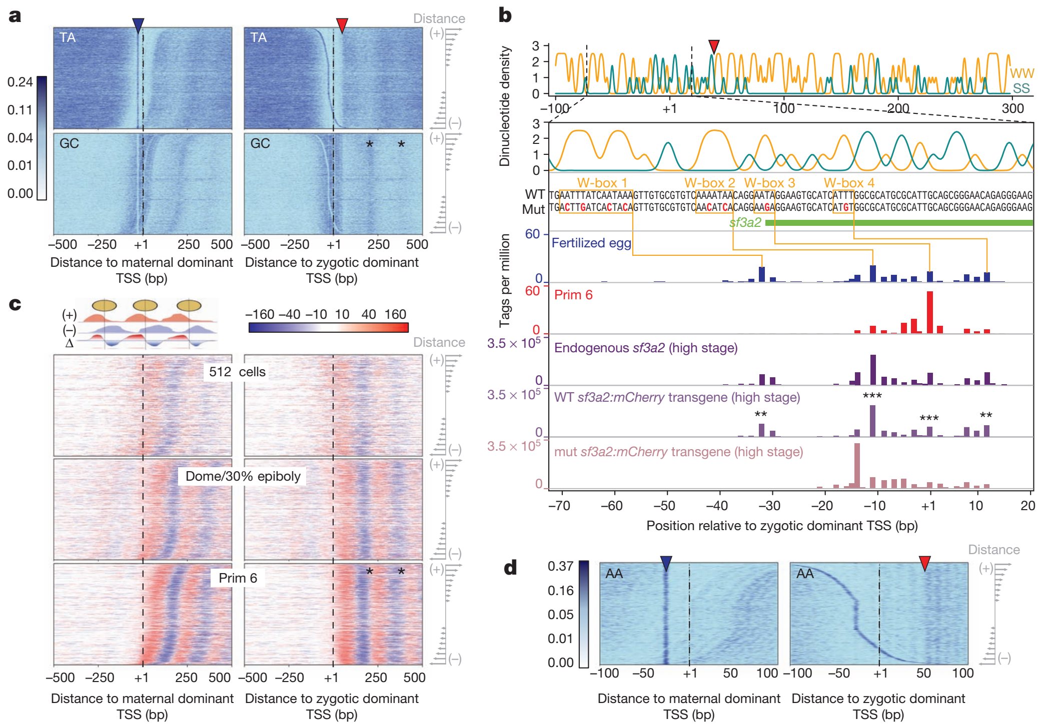


Figure 3 | Transition from maternal W-box-motif-dependent to zygotic nucleosome-positioning-signal-related transcription initiation is pervasive. **a**, Dinucleotide density at 8,369 constitutively expressed promoters sorted by the distance between the maternal and zygotic dominant TSS. Promoters were centred at either the maternal (left) or the zygotic (right) dominant TSS. Blue arrowhead indicates position of the maternal TSS-associated W-box; red arrowhead indicates the SS–WW boundary ~50 bp downstream of zygotic TSS; asterisks indicate GC enrichment in the internucleosomal region. **b**, Predicted maternal and zygotic codes in the *sf3a2* promoter. Dinucleotide density and sequence of the wild-type (WT) and mutated (mut) *sf3a2* promoter are shown on top. TSSs detected by CAGE in wild-type zebrafish in the maternal and zygotic stages are shown in blue and red, respectively. The W-boxes associated with maternal TSSs are marked in orange, and the introduced point mutations disrupting them in red. SL-CAGE TSSs in stable transgenic lines for endogenous *sf3a2*, the wild-type *sf3a2* transgene and mutant *sf3a2* transgene are shown in different shades of purple. $^{**}P < 0.01$, $^{***}P < 0.001$, one-tailed Welch's two sample *t*-test; mutant, $n = 4$; wild type, $n = 3$. **c**, Subtracted H3K4me3 coverage (Δ) of reads mapping to the plus (+) and minus (−) strand (schematic on top) in three developmental stages at the same set of promoters as in **a**. **d**, Density of AA dinucleotide in a ±100 bp region for promoters in **a**.

containing only one broad TC (Extended Data Fig. 5a–c). Constitutively used promoters changed their shape accordingly, from a single or multiple sharp TSS configuration in the maternal stages, to a broad configuration in the zygotic stages (Extended Data Fig. 5d, e). Thus, the switch between the maternal and zygotic TSS is accompanied by a global change in the promoter architecture within the same region.

To validate functionally the observed TSS selection grammars, we identified the W-boxes and dinucleotide frequency patterns in a constitutively active promoter (*sf3a2*) and mutated all W-boxes associated with maternal TSSs (Fig. 3b) for analysis in transgenic zebrafish. Fluorescence reporter activity and 5′ rapid amplification of cDNA ends (RACE) assays demonstrated that removal of all W-boxes did not influence zygotic transgene activity or zygotic TSS selection (Extended Data Fig. 6), confirming W-box-independent promoter usage in the embryo. To validate W-box-dependent TSS selection in the oocyte, we analysed maternal TSS selection in early F1 generation embryos from stable transgenic lines with wild-type or mutated variants of the *sf3a2* promoter. We developed a novel method (single-locus CAGE (SL-CAGE); Supplementary Table 2) for the detection and relative quantification of TSS usage at 1 bp resolution within the targeted promoter. TSS usage patterns of the wild-type *sf3a2* transgene promoter in early embryos of several transgenic lines were highly reproducible and perfectly correlated with maternal TSS usage of the endogenous *sf3a2* gene as well as with that seen by CAGE (Fig. 3b and Extended Data Fig. 7). The removal of W-boxes severely reduced the use of associated downstream positions as TSSs in the mutant transgenic lines compared to the wild-type lines and led to an aberrant TSS usage pattern (Fig. 3b and Extended Data Fig. 7d; $P < 0.01$), confirming that the selection of these TSSs depends on a W-box signal in the oocyte. These results indicate two independent TSS selection mechanisms used by the oocyte and the embryo within a single promoter.

To address the relationship between stage-specific TSS selection and chromatin configuration, we analysed the positioning of H3K4me3- and H2A.Z-containing nucleosomes at core promoters by chromatin immunoprecipitation followed by sequencing (ChIP-seq). It was previously shown that H3K4me3 marking on promoter-associated nucleosomes precedes gene transcription during zebrafish genome activation[5,6]. Our data revealed precise positioning of the first downstream (+1) nucleosome ~50 bp from the preferred zygotic TSS, but no fixed spacing to the maternal TSS for all constitutively active promoters in the zygotic (prim 6) stage (Fig. 3c and Extended Data Fig. 3d). We observed a less sharp nucleosome alignment to the zygotic TSS in earlier stages, including the 512-cells stage, which precedes the onset of zygotic transcription. By contrast, no alignment of H3K4me3-marked nucleosomes to the maternal TSS was detected in any stage (Fig. 3c). These results revealed a positional interdependency between the zygotic TSS and the +1 nucleosome in the embryo as a feature of zygotic TSS selection grammar, independent of the W-box motif-guided TSS selection in the oocyte. Promoter-associated nucleosome alignment corresponded with alternating WW–SS patterns downstream of the zygotic TSS (asterisks in Fig. 3a, c), providing an internucleosomal position signal. At a higher resolution, the nucleosome-occupied DNA downstream of the zygotic TSS showed a 10 bp periodicity in AA and TT dinucleotide enrichment (Fig. 3d and Extended Data Fig. 3e), previously identified as an intranuclesomal positioning signal[16]. The strong association of the zygotic, but not the maternal, TSS with these nucleosome positioning signals

suggests that TSS selection in a vertebrate oocyte is independent of inter- and intranucleosomal DNA signals.

Recent efforts to identify sequence-based signals for nucleosome positioning[17,18] and dynamic nucleosome organization at promoters[19,20] highlight the epigenetic and chromatin mechanisms[21,22] that, together with the DNA sequence, direct transcription initiation. The association of nucleosome positioning signals with zygotic promoter activity described here raises the question of whether promoter-associated nucleosome positioning contributes to regulation of positioning of transcription initiation, or is merely a consequence of transcription at the predefined position. To investigate this relationship we analysed the DNA sequence underlying +1 nucleosome positioning in the transcriptionally silent pre- (512 cells) and active post-MBT (prim 6) stages (Fig. 4a and Extended Data Fig. 8). In the pre-MBT stage, H3K4me3-marked nucleosomes occupied a CG/GC enriched region and centred at the peak of highest CG/GC enrichment, often directly overlapping the TSSs of the maternal transcripts, supporting the idea that H3K4me3 initially appears at CpG islands before transcription[23]. In the post-MBT stage, the +1 nucleosome was positioned just downstream of the SS–WW enrichment boundary at ~50 bp from the zygotic TSS, occupying a WW-enriched region, with a small local GC/CG peak at the nucleosome midpoint (Fig. 4a and Extended Data Fig. 8b). Additional downstream nucleosomes followed a similar pattern of WW-enriched bound DNA alternating with internucleosomal SS enrichment. Local GC/CG enrichment at the nucleosome midpoints is in accordance with previously described nucleosome positioning preferences[16]; however, the additional sequence preference complexity and its relation to TSS in different developmental stages were not reported so far. The results show that the initial positioning of promoter-associated nucleosomes, which correlates with a broad internucleosomal phasing pattern, changes in later stages to a final precise positioning, which correlates tightly with the zygotic TSS and intranucleosomal phasing patterns, suggesting interdependence of the final nucleosome positioning and transcription.

To test this, we ranked throughout-active genes by the timing of onset of their zygotic transcription and analysed their H3K4me3-marked nucleosome positioning patterns (Fig. 4b). No association between the timing of transcription activation and the precision of nucleosome positioning was found, arguing against transcriptionally aided nucleosome readjustment and instead suggesting a pre-transcriptional process for repositioning nucleosomes to their final position, in agreement with transcription-independent positioning of nucleosomes at promoters in human cells[24]. Consistently, H3K4me3 ChIP-seq in TATA-box-binding protein (Tbp) knockdown embryos (Extended Data Fig. 9) showed no change in the overall H3K4me3 recruitment and nucleosome positioning at Tbp-dependent genes (Fig. 4c), demonstrating that H3K4me3-marked nucleosome positioning at these genes does not require Tbp-dependent recruitment of the transcription initiation machinery or active transcription.

The absence of a nucleosome-positioning sequence signature, as well as of precise nucleosome positioning at promoters with a canonical TATA-box in other systems[20,25], together with narrow TSS peaks, argues in favour of the W-box as the overriding determinant of maternal TSS selection. The similarity of the W-box to the TATA-box suggests that transcription initiation in the oocyte may be mediated by the oocyte-enriched transcription nucleating factor Tbpl2 (also known as Tbp2)[26,27]. Conversely, early zygotic grammar prefers the TSS position to be at a fixed range from the precisely positioned +1 nucleosome, suggesting a mechanism in which the initiation complex chooses initiator-like sequences within a ‘catchment area’ determined by the nucleosome position (Fig. 4d). This model is compatible with motif-independent TFIID recruitment by H3K4me3–TAF3 interactions[28] and emphasizes the interdependence of nucleosome configuration at promoters with promoter type and physiological state in vertebrates[20,25] and yeast[19,29].

Different TSS selection grammars deployed at separate promoters have been associated with different types of genes[19,20], and a handful of promoters were shown to switch between TATA-dependent and -independent initiation[30]. We show for the first time, to our knowledge,

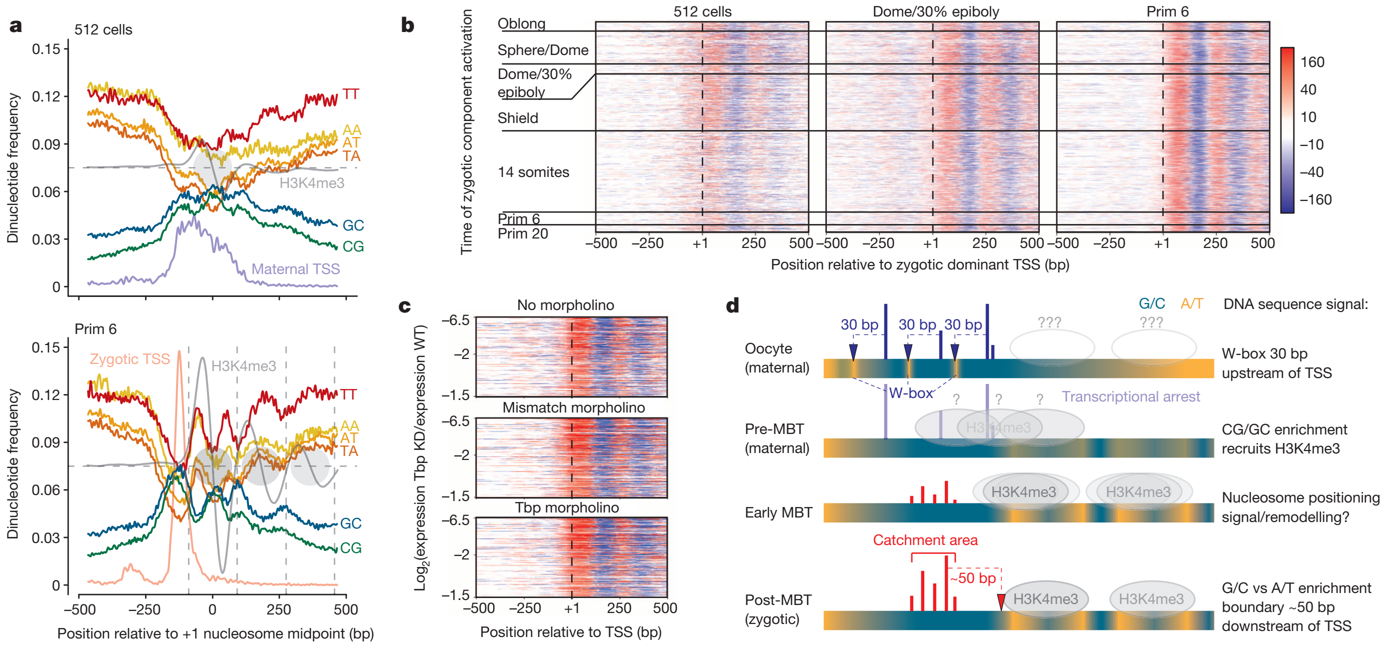


Figure 4 | H3K4me3-marked nucleosome positioning reveals dynamic changes in underlying sequence signature and relation to TSS during maternal to zygotic transition. **a**, Frequency of dinucleotides centred on +1 nucleosome of constitutively active promoters in maternal (512 cells) and zygotic (prim 6) stages. Centres of nucleosomes were estimated from subtracted H3K4me3 coverage (grey). Density of maternal and zygotic TSSs is shown in light blue and light red, respectively. **b**, H3K4me3 signal at promoters of constitutively present transcripts sorted by the time of activation of their zygotic component. Horizontal lines separate groups of promoters that activate the zygotic component at a denoted developmental stage. **c**, H3K4me3 signal at Tbp-dependent promoters in non-injected embryos (top), embryos injected with mismatch morpholino (middle) or Tbp-targeting morpholino (bottom), sorted by expression fold change between Tbp knockdown (KD) and wild-type (WT) embryos. **d**, Summary of transcription initiation, TSS configuration and nucleosome positioning dynamics throughout maternal to zygotic transition.

that the two grammars co-exist in close proximity or in physical overlap genome-wide, and are differentially used at thousands of promoters active in both the oocyte and the embryo. The multiple layers of information embedded in the same short sequence, each representing a different aspect of a complex regulation, are part of the reason why promoter codes have been so difficult to detect. Our findings on overlapping promoter grammars have implications for future analyses of promoter content and function.

METHODS SUMMARY

TSSs defined by CAGE were clustered into TCs, and TCs into promoter regions, allowing a maximal distance of 20 and 100 bp, respectively. Dinucleotide density was calculated by applying two-dimensional Gaussian kernel smoothing on a dinucleotide occurrence matrix. TSS usage in the *sf3a2* promoter was analysed in transgene-injected embryos by 5′ RACE and in stable transgenic zebrafish lines by SL-CAGE. ChIP was carried out on MNase-treated chromatin using the ChIP-IT Express Enzymatic kit (Active Motif), and libraries were sequenced for 36 or 56 bp on the HiSeq 2000 system. Minus-strand reads coverage was subtracted from plus-strand reads coverage for visualization and nucleosome midpoint estimation. Tbp was knocked down by injecting one-cell-stage zebrafish embyos with the *tbp* morpholino.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 February; accepted 23 December 2013.

Published online 16 February; corrected online 19 March 2014 (see full-text HTML version for details).

1. D'Alessio, J. A., Wright, K. J. & Tjian, R. Shifting players and paradigms in cell-specific transcription. *Mol. Cell* **36,** 924–931 (2009).
2. Kadonaga, J. T. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip. Rev. Dev. Biol.* **1,** 40–51 (2012).
3. Mathavan, S. *et al.* Transcriptome analysis of zebrafish embryogenesis using microarrays. *PLoS Genet.* **1,** e29 (2005).
4. Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts. *Development* **136,** 3033–3042 (2009).
5. Vastenhouw, N. L. *et al.* Chromatin signature of embryonic pluripotency is established during genome activation. *Nature* **464,** 922–926 (2010).
6. Lindeman, L. C. *et al.* Prepatterning of developmental gene expression by modified histones before zygotic genome activation. *Dev. Cell* **21,** 993–1004 (2011).
7. Kane, D. A. & Kimmel, C. B. The zebrafish midblastula transition. *Development* **119,** 447–456 (1993).
8. Shiraki, T. *et al.* Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc. Natl Acad. Sci. USA* **100,** 15776–15781 (2003).
9. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823–837 (2007).
10. Carninci, P. *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genet.* **38,** 626–635 (2006).
11. Lenhard, B., Sandelin, A. & Carninci, P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Rev. Genet.* **13,** 233–245 (2012).
12. Venters, B. J. & Pugh, B. F. Genomic organization of human transcription initiation complexes. *Nature* **502,** 53–58 (2013).
13. Nepal, C. *et al.* Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. *Genome Res.* **23,** 1938–1950 (2013).
14. Giraldez, A. J. *et al.* Zebrafish miR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312,** 75–79 (2006).
15. Ponjavic, J. *et al.* Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. *Genome Biol.* **7,** R78 (2006).
16. Segal, E. *et al.* A genomic code for nucleosome positioning. *Nature* **442,** 772–778 (2006).
17. Ioshikhes, I., Hosid, S. & Pugh, B. F. Variety of genomic DNA patterns for nucleosome positioning. *Genome Res.* **21,** 1863–1871 (2011).
18. Segal, E. & Widom, J. What controls nucleosome positions? *Trends Genet.* **25,** 335–343 (2009).
19. Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483,** 295–301 (2012).
20. Rach, E. A. *et al.* Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. *PLoS Genet.* **7,** e1001274 (2011).
21. Cairns, B. R. The logic of chromatin architecture and remodelling at promoters. *Nature* **461,** 193–198 (2009).
22. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25,** 1010–1022 (2011).
23. Thomson, J. P. *et al.* CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464,** 1082–1086 (2010).
24. Fenouil, R. *et al.* CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res.* **22,** 2399–2408 (2012).
25. Nozaki, T. *et al.* Tight associations between transcription promoter type and epigenetic variation in histone positioning and modification. *BMC Genomics* **12,** 416 (2011).
26. Bártfai, R. *et al.* TBP2, a vertebrate-specific member of the TBP family, is required in embryonic development of zebrafish. *Curr. Biol.* **14,** 593–598 (2004).
27. Akhtar, W. & Veenstra, G. TBP2 is a substitute for TBP in *Xenopus* oocyte transcription. *BMC Biol.* **7,** 45 (2009).
28. Lauberth, S. M. *et al.* H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* **152,** 1021–1036 (2013).
29. Zaugg, J. B. & Luscombe, N. M. A genomic model of condition-specific nucleosome behavior explains transcriptional activity in yeast. *Genome Res.* **22,** 84–94 (2012).
30. Davis, W. Jr & Schultz, R. M. Developmental change in TATA-box utilization during preimplantation mouse development. *Dev. Biol.* **218,** 275–283 (2000).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors are grateful to L. Tora, E. Kenyon, G. Chalancon and J. C.-y. Chen for comments on the manuscript, to M. Reischl for Zebrafish Miner software, and to L. O'Neill for technical advice. V.H., C.N., C.Pr., A.A. and X.D. were supported by grants from the Norwegian Research Council (YFF) and the Bergen Research Foundation, awarded to B.L. F.M., U.S. and B.L. acknowledge support from European Union (EU) FP6 integrated project EuTRACC and FP7 integrated project ZF Health. B.L. was additionally supported by the Medical Research Council UK, F.M. and P.C. by the EU FP7 project Dopaminet and U.S. by the EU FP6 project NeuroXSys. C.Pl., A.M.S. and P.C. were supported by a Research Grant from MEXT to RIKEN CLST. (RIKEN Omics Science Center ceased to exist on 1 April 2013 owing to reorganization of RIKEN.)

Author Contributions B.L., F.M., P.C. and V.H. conceived the study. N.L., Y.H., J.G. and M.F. performed experiments. O.A. and W.F.J.v.I. performed sequencing. V.H., C.Pr., C.N., X.D. and A.A. performed computational analyses. C.Pl. and A.M.S. developed and performed SL-CAGE with input from P.C. V.H., B.L. and F.M. analysed the data and wrote the manuscript with input from P.C., W.F.J.v.I. and U.S.

Author Information High-throughput sequencing data have been deposited at the NCBI Sequence Read Archive under accession numbers SRA097279 and SRA104816. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.L. (b.lenhard@imperial.ac.uk), F.M. (f.mueller@bham.ac.uk) or P.C. (carninci@riken.jp).

METHODS

CAGE tags mapping and CTSS calling. Sequenced CAGE tags (27 bp) as previously described[13] were mapped to a reference zebrafish genome (Zv9/danRer7 assembly) using Bowtie[31] with default parameters allowing up to two mismatches and keeping only uniquely mapped reads. An additional G nucleotide, which is often attached to the 5′ end of the tag by the template-free activity of the reverse transcriptase in the cDNA preparation step of the CAGE protocol[32], was removed in cases in which it did not map to the genome. All unique 5′ ends of tags were considered as CAGE tag-defined TSSs (CTSSs) and the number of tags supporting each CTSS was counted. Raw tag count was normalized to a referent power-law distribution based on a total of 10^6 tags and $\alpha = -1.25$, as described previously[33], resulting in normalized tags per million (tpm). All analyses were done in the R statistical computing environment[34] (http://www.R-project.org/) using Bioconductor[35] (http://www.bioconductor.org/) software packages and custom scripts.

CTSS clustering into TCs and promoter regions. CTSSs supported by at least 1 tpm in at least one of the 12 developmental stages were clustered at two levels. First, TCs were created for each stage individually using a simple distance-based approach with a maximum allowed distance of 20 bp between two neighbouring CTSSs. Next, for each TC we calculated a cumulative distribution of the CAGE signal and determined the positions of the 10th and 90th percentile to obtain more robust boundaries of a TC. TCs across all developmental stages within 100 bp of each other were aggregated into a single promoter region. Only promoter regions supported by at least 5 tpm in at least one developmental stage were used in further analyses.

Expression profiling. Expression profiling was done at two levels: individual CTSSs, and entire promoter regions. To minimize the noise from weakly supported CTSSs, we selected only CTSSs with at least 5 tpm in at least one developmental stage. Normalized tpm values across 12 developmental stages for each CTSS (or promoter region) were divided by their standard deviation to obtain scaled expression measures. A self-organizing map[36] (SOM) unsupervised learning algorithm was applied to distribute CTSSs (or promoter regions) across $5 \times 5 = 25$ expression profiles.

Dinucleotide pattern analysis. To visualize dinucleotide composition patterns of sequences flanking TSSs, we first created an occurrence matrix ($n \times m$; where n = number of sequences and m = length of sequences) for each individual dinucleotide, by placing 1 if the given dinucleotide is present at a given position or 0 if it is not. Values in the matrix were then smoothened: at each position in the matrix the weighted average dinucleotide occurrence was calculated by taking into account surrounding positions. Weights of the surrounding positions were assigned by centring a two-dimensional Gaussian kernel with bandwidth = 3 (in both dimensions) at the central position. The matrix of smoothened values (densities) was visualized using different shades of blue in a map-like representation. Extended Data Fig. 10 illustrates how the calculation and visualization was done.

TATA-box motif analysis. A TATA-box PWM was obtained from the JASPAR database[37] (http://jaspar.genereg.net/) and used to scan the region −35 bp to −22 bp upstream of the TSS (expected position for a TATA-box was determined as previously described[15]). For each promoter sequence a maximal detected match (%) to the TATA-box PWM was reported for the maternal and zygotic dominant TSS separately. Distribution of obtained values across all promoters was visualized by histograms. In addition, the frequency of the top ten most abundant pentamers found in the scanned sequences was shown. *De novo* motif discovery was performed on a set of 14 bp long sequences spanning the region from −35 bp to −22 bp upstream of the TSS using MEME[38] (http://meme.sdsc.edu/) with default parameters. Only motifs with an E value ≤ 0.01 were selected as significant.

***sf3a2* promoter reporter constructs.** The region spanning 500 bp upstream and 200 bp downstream of the dominant zygotic TSS in the *sf3a2* promoter was chosen for validation of the TSS selection grammar. The *sf3a2* promoter carries both maternal and zygotic promoter determinants and exhibits TSS shifting. The selected sequence, which ends within the first intron of *sf3a2*, was fused to a sequence containing the 3′ end of the zebrafish *txnipa* first intron and splice acceptor fused to an mCherry reporter (Extended Data Fig. 6a). Genomic DNA of AB* zebrafish strain was used for PCR amplification using Advantage HD DNA Polymerase Mix (Clontech). The amplified fragments were cloned into a pDB896 vector (provided by D. Balciunas and subsequently modified by replacing the *γ-crystallin:gfp* with *mCherry*) using In-Fusion PCR Cloning System (Clontech) following the manufacturer's instructions. The expression cassette is flanked by Tol2 transposon arms for Tol2 transgenesis[39]. A polymorphic nucleotide (G>T) was identified in the promoter sequence at chr2:58,656,711 (Ensembl, Zebrafish Assembly Zv9/danRer7). The wild-type promoter reporter vector was used as a template for in site-directed mutagenesis PCR to introduce the mutation in the W-boxes as indicated in Fig. 3b. The resulting mutated PCR fragment was cloned in the same reporter vector using the In-Fusion PCR Cloning System. Sequences of all primers are provided in Extended Data Fig. 6b.

Microinjection and transgene expression analysis. 1.5–2 nl of injection solution containing 20 ng μl^{-1} reporter plasmid DNA and 15 ng μl^{-1} Tol2 transposase mRNA, supplemented with 0.1% Phenol red (injection marker), was injected into zebrafish eggs within 10–15 min after fertilization. In the automated imaging and expression analysis experiments, *ecfp* mRNA was added (30 ng μl^{-1}). mCherry reporter activity was measured in the prim 20 stage, in both embryos injected with wild-type and mutated *sf3a2* promoter constructs by automated imaging as described previously[40]. Embryo images were analysed with Zebrafish Miner software[40]. The level of reporter expression was measured as pixel intensity value and normalized to the intensity of the ECFP signal (injection control) and averaged for all embryos in the experiment. In addition a percentage of expressing embryos was calculated from the total number of ECFP-positive embryos with an ECFP signal equal or above the detection threshold of the Zebrafish Miner software. Embryo images were wrapped onto reference embryo shape and overlaid by summing pixel intensity values.

5′ RACE. 5′ RACE was performed with a FirstChoice RLM-RACE Kit (Life Technologies) following the manufacturer's protocol. Total RNA was isolated in the prim 6 stage from ~100 phenotypically normal looking and reporter-gene-expressing zebrafish embryos injected with either the wild-type or mutated *sf3a2* promoter reporter construct, using TRIZOL (Life Technologies) following the manufacturer's instructions. The PCR products of the expected size from the nested (inner) PCR reaction were purified from agarose gel and sequenced to identify the TSS. To demonstrate that the generated 5′ RACE products are specific to the 5′ ends of de-capped RNA, a 'minus tobacco acid pyrophosphatase' (TAP)-treated sample was put through adaptor ligation, reverse transcription and PCR. Sequences of primers used in 5′ RACE are provided in Extended Data Fig. 6f.

Transgenic zebrafish lines. Transgenic zebrafish lines with the *sf3a2* promoter (wild-type and mutated) reporter constructs were generated by microinjection of the corresponding construct into zebrafish zygotes as described earlier. The reporter-positive (*mCherry*) embryos were grown to adulthood and germline-transmitting female individuals were identified by crossing to wild-type zebrafish and selecting for the presence of reporter-expressing offspring. Transgene expression and TSS usage was analysed in F1 generation embryos. Experiments were carried out under licence by the Home Office Licence Number 40/3681 and PPL 40/3131.

Quantification of the reporter mRNA levels by qPCR. RNA from reporter-expressing embryos in the high/sphere stage was isolated using GeneElute Total RNA extraction kit (Sigma-Aldrich), following the manufacturer's instructions. qPCR was performed using the SYBR Green detection method on 7900HT Fast Real-Time PCR System (Applied Biosystems). Two primer pairs were used for both the *mCherry* reporter and the endogenous *sf3a2* gene (normalization control). Technical triplicates were run for each primer pair. The C_t values were determined by the SDS v.2.4 software (Applied Biosystems), using a manual threshold of 0.2 and automatic base line. Expression levels of the transgene were calculated relative to the endogenous *sf3a2* in the same sample, using the average C_t values across technical triplicates and both primer pairs. The sequence of the primers used in qPCR is provided in Extended Data Fig. 7c.

SL-CAGE. We introduced a novel method for quantitative high-resolution detection of TSSs and their usage within a targeted promoter, called single-locus deep CAGE. The method combines generation of 5′ complete cDNAs transcribed from capped mRNAs as described in the CAGE protocol[8] with the amplification of targeted cDNAs using gene-specific primers and subsequent high-throughput paired-end sequencing of the single locus (typically single promoter region) based library. Supplementary Table 2 describes the main steps of the protocol and provides sequences of all primers used in different steps. Paired-end sequenced reads (34 bp + 35 bp) were mapped to either the spliced sequence of the *sf3a2:mCherry* transgene or endogenous *sf3a2*, and TSS usage was reconstructed as described earlier for CAGE tags.

ChIP. ChIP experiments were carried out using the ChIP-IT Express Enzymatic kit (Active Motif) in line with the manufacturer's instructions. Chromatin was prepared using ~5,000 and ~3,000 embryos in the 512-cell and oblong stages, respectively. Embryos were dechorionated enzymatically using pronase and fixed in 1.85% formaldehyde in Hanks media for 20 min at room temperature (23 °C). The embryos were washed once with PBS and the fixation was stopped by incubating in 1× glycine for 10 min at room temperature followed by three washes with ice-cold PBS. Embryos were resuspended in 1 ml ice-cold lysis buffer, incubated on ice for 20 min, transferred to a pre-cooled dounce homogenizer and dounced by 10 strokes. Nuclei were collected by centrifugation, resuspended in 200 μl digestion buffer and incubated at 37 °C for 5 min. Chromatin was sheared by adding 10 μl of enzymatic shearing cocktail working stock (200 U ml^{-1}) and incubating for 10 min at 37 °C. Shearing efficiency was checked by gel electrophoresis according to manufacturer's instructions. The reaction was stopped by adding 5 μl ice-cold 0.5 M EDTA and incubating on ice for 10 min and sheared chromatin was cleared by centrifugation. For ChIP reactions 70 μl of sheared chromatin were mixed with

25 μl Protein G magnetic beads, 20 μl ChIP buffer 1, 1 μl protein inhibitor cocktail, 4 μg of anti-H3K4me3 (Abcam ab8580) or anti-H2A.Z (Abcam ab4174) antibody or an equivalent volume of water (no antibody control), respectively, and water to a final volume of 200 μl. ChIP was performed in duplicates for each stage. ChIP reactions were then incubated overnight at 4 °C while rotating. Magnetic beads were washed and incubated in elution buffer. After addition of reverse crosslinking buffer samples were decrosslinked for 4 h at 65 °C. Samples were treated with Proteinase K and RNase A and purified using pheonol chloroform extraction.

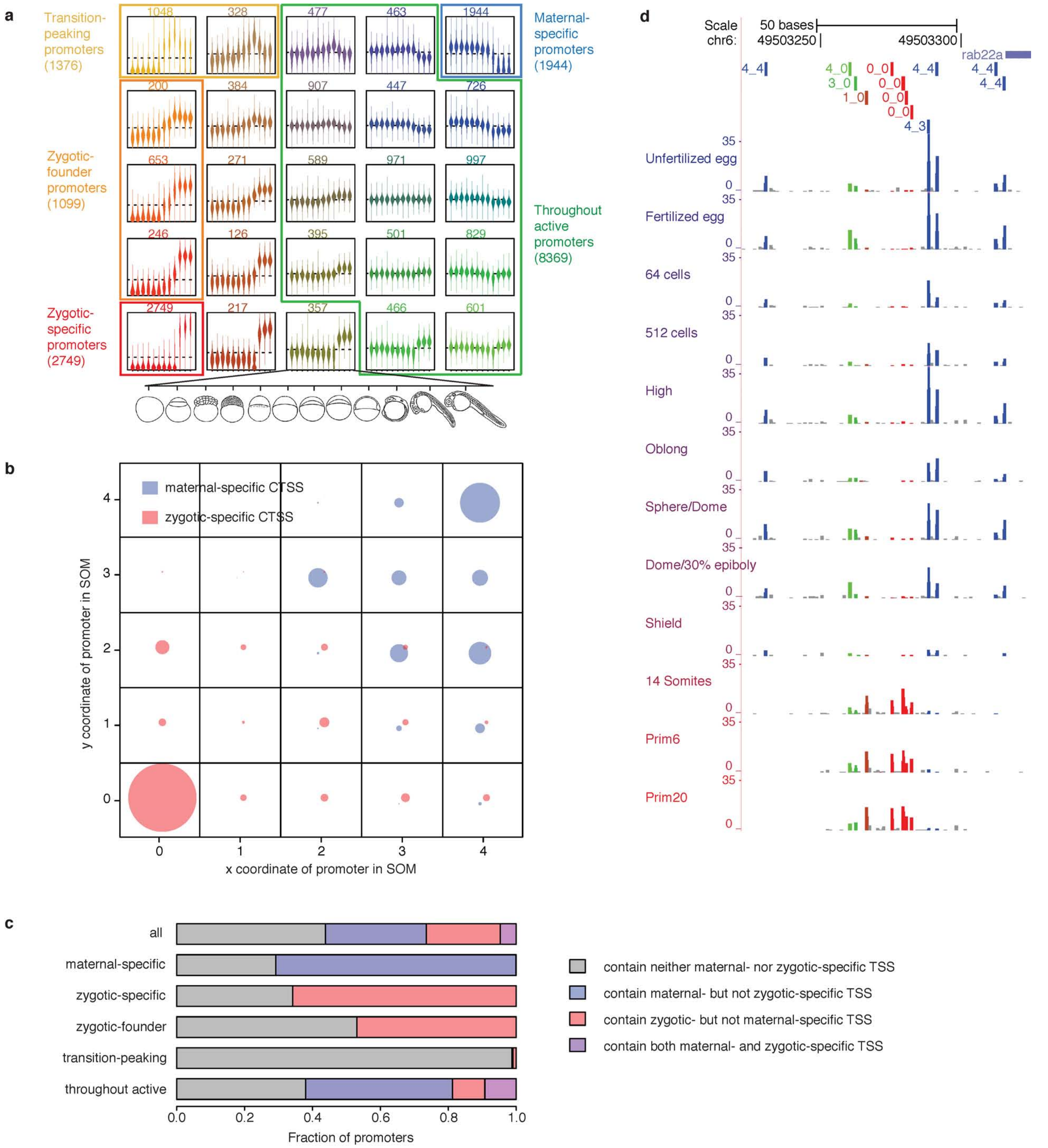
Tbp knockdown. One-cell-stage embryos were injected with either *β-actin:yfp*[41] (1.7 nl, 42 pg nl^{-1}) or 1.3 kb upstream of the *ntl* TSS *ntl:yfp*[42] (1.7 nl, 68 pg nl^{-1}) constructs. The injected embryos were then split into four groups. One was kept as a non-injected control and the other three groups were further injected with one of the two *tbp*-targeting morpholinos (1.7 nl, 2.5 mM) or with a mismatch morpholino described previously[43]. All embryos were kept in E3 medium at 28.5 °C until the non-injected group reached the 30% epiboly stage (4.7 hpf) and were then analysed under a fluorescence stereoscope (Nikon SMZ1500). Arrest of epiboly movements, loss of *β-actin:yfp* (Tbp-dependent) and retention of *ntl:yfp* (Tbp-independent) reporter activities[42] were used as markers for assessing and sorting Tbp morphants for ChIP analysis. Approximately 1,500 non-injected, 1,500 mismatch-morpholino-injected embryos, 1,200 *tbp mo1* morphants and 1,000 *tbp mo2* morphants were used for ChIP, as described earlier. We used a previously published set of genes downregulated in zebrafish Tbp morphants[43]. Genes with log fold change ≤ -1.5 were selected as Tbp dependent and were aligned with respect to the dominant TSS of the nearest promoter detected by CAGE for H3K4me3 ChIP-seq signal visualization.

Deep sequencing of chromatin DNA. ChIP-seq was performed as described before[44]. In brief, 10 ng of ChIP DNA was end-repaired, ligated to single read adaptors, size selected and amplified for 18 cycles according to Illumina's ChIP-seq protocol. Cluster generation was performed according to the Illumina Cluster Reagents preparation protocol (http://www.illumina.com/). Samples were sequenced for 36 bp or 56 bp (*tbp* morphants and controls) on the HiSeq 2000 system.

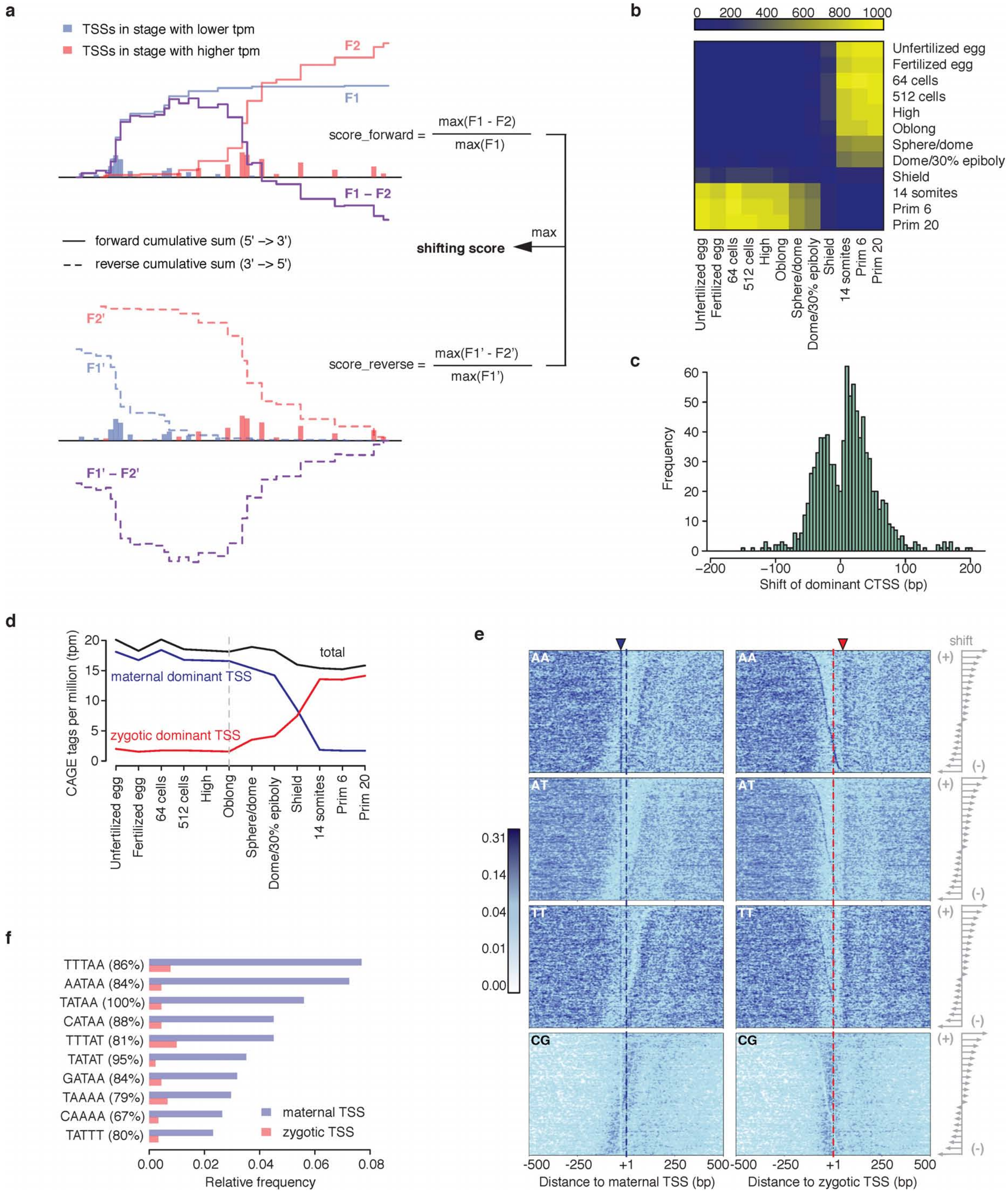
ChIP-seq data analysis. Sequenced reads were mapped to the reference zebrafish genome (Zv9/danRer7 assembly) using Bowtie[31] with default parameters allowing up to two mismatches and keeping only uniquely mapped reads. Coverage was calculated for plus and minus strands separately using unextended reads and taking maximum 20 reads mapping to exactly the same position. Minus-strand coverage was subtracted from plus-strand coverage to obtain subtracted coverage, which was used for visualization and nucleosome midpoint estimation. Significantly enriched regions (peaks) were detected using model-based analysis for ChIP-seq (MACS)[45] (http://liulab.dfci.harvard.edu/MACS/) with default parameters. Midpoints of nucleosomes within significantly enriched regions (FDR $\leq$ 0.01) were estimated from subtracted coverage and the nearest CAGE signal was used to determine strand specificity and relative position of the first downstream nucleosome.

Data accession. Processed data are available for download at http://promshift.genereg.net/zebrafish/. Tracks can be visualized as annotated custom tracks in the UCSC Genome Browser using the following URLs: http://promshift.genereg.net/zebrafish/CAGE_and_nucleosome_tracks.txt and http://promshift.genereg.net/zebrafish/Transgenic_lines_sf3a2_SL-CAGE_tracks.txt.

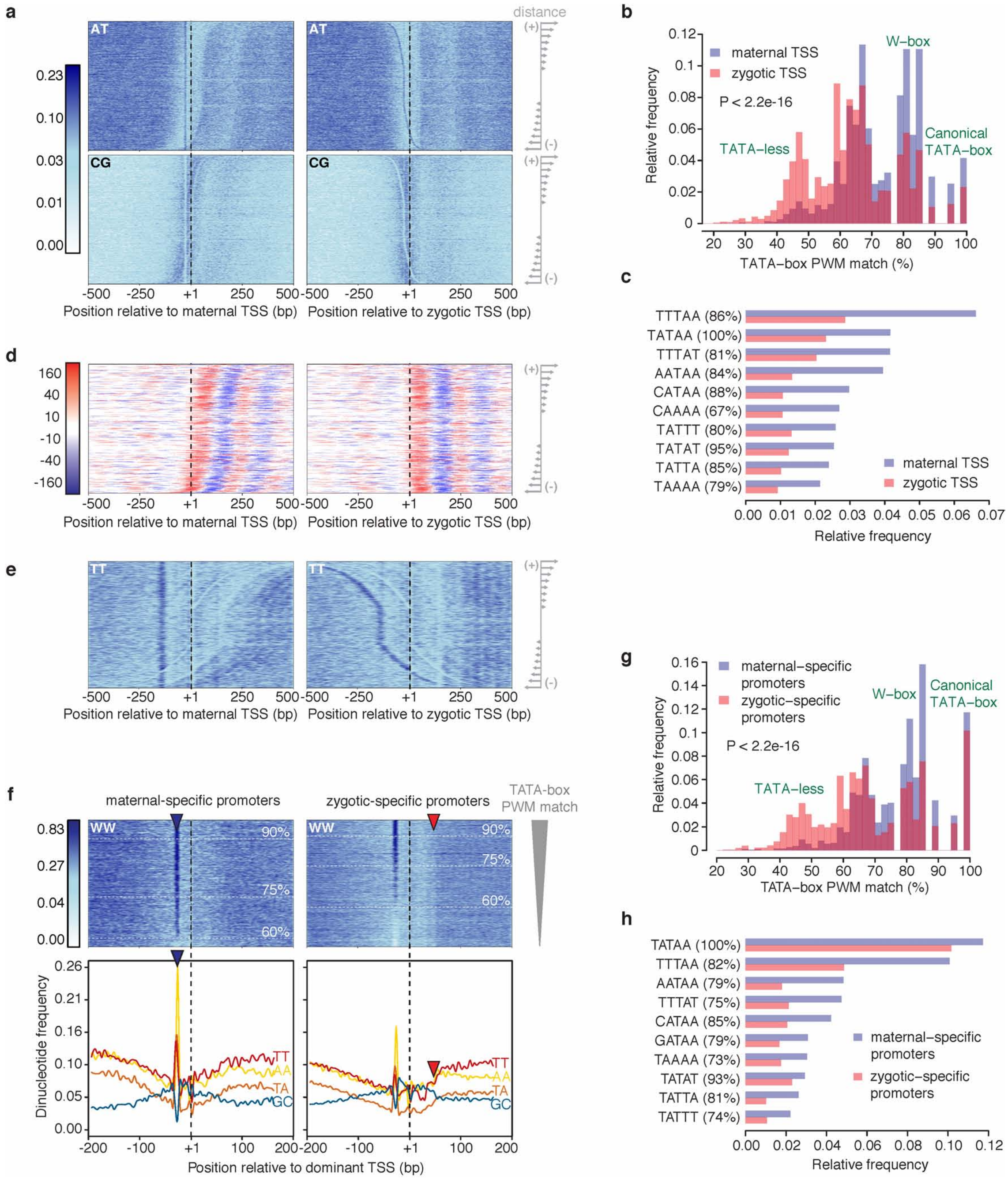
31. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
32. Kodzius, R. *et al.* CAGE: cap analysis of gene expression. *Nature Methods* **3,** 211–222 (2006).
33. Balwierz, P. J. *et al.* Methods for analyzing deep sequencing expression data: constructing the human and mouse promoterome with deepCAGE data. *Genome Biol.* **10,** R79 (2009).
34. The R Development Core Team. *R: A Language and Environment for Statistical Computing* 1–3079 (R Foundation for Statistical Computing, 2013).
35. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5,** R80 (2004).
36. Törönen, P., Kolehmainen, M., Wong, G. & Castren, E. Analysis of gene expression data using self-organizing maps. *FEBS Lett.* **451,** 142–146 (1999).
37. Portales-Casamar, E. *et al.* JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **38,** D105–D110 (2010).
38. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Sec. Int. Conf. Intell. Syst. Mol. Biol.* **2,** 28–36 (1994).
39. Abe, G., Suster, M. L. & Kawakami, K. in *The Zebrafish: Genetics, Genomics and Informatics* (eds Detrich, H. W., Westerfield, M. & Zon, L. I.) Vol. 104, 23–49 (Academic, 2011).
40. Gehrig, J. *et al.* Automated high-throughput mapping of promoter-enhancer interactions in zebrafish embryos. *Nature Methods* **6,** 911–916 (2009).
41. Higashijima, S., Okamoto, H., Ueno, N., Hotta, Y. & Eguchi, G. High-frequency generation of transgenic zebrafish which reliably express GFP in whole muscles or the whole body by using promoters of zebrafish origin. *Dev. Biol.* **192,** 289–299 (1997).
42. Ferg, M. *Large Scale- and Functional Analysis for the Requirement of TBP-Function in Early Zebrafish Development.* PhD thesis, Heidelberg Univ. (2008).
43. Ferg, M. *et al.* The TATA-binding protein regulates maternal mRNA degradation and differential zygotic transcription in zebrafish. *EMBO J.* **26,** 3945–3956 (2007).
44. Soler, E. *et al.* A systems approach to analyze transcription factors in mammalian cells. *Methods* **53,** 151–162 (2011).
45. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).



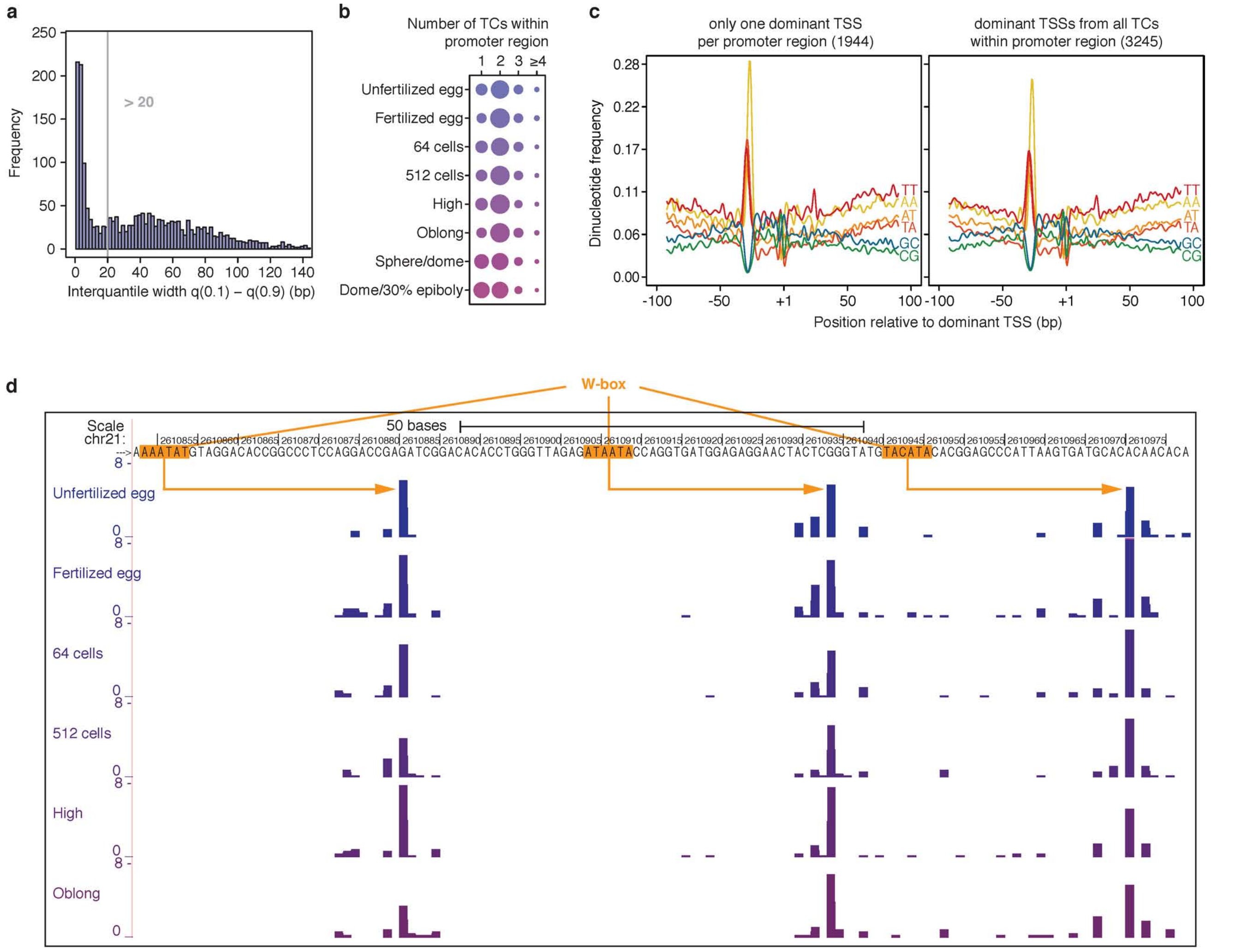
Extended Data Figure 1 | Promoters contain TSSs with very heterogeneous developmental expression profiles. **a**, Developmental expression profiles of CAGE-defined promoters obtained by self-organizing map (SOM) clustering. Each box represents one SOM cluster, with a series of beanplots showing distribution of scaled expression (logarithm of normalized number of CAGE tags per million) at different time points for all promoters belonging to that cluster (number of promoters denoted above the box). Different classes of promoters marked in different colours were used in all downstream analyses. **b**, Projection of maternal- and zygotic-specific CTSSs (from Fig. 1b) onto the SOM classes of promoters shown in **a**. **c**, Fraction of promoters containing either maternal-specific (cluster 4,4 in Fig. 1b) or zygotic-specific (cluster 0,0 in Fig. 1b) CTSS, or both, for different classes of promoters established in **a**. **d**, Example of a throughout-active promoter that contains a very heterogeneous population of CTSSs (colour coding of CTSSs as in Fig. 1b: maternal-specific shown in blue, throughout-expressed shown in green, and zygotic-specific shown in red).



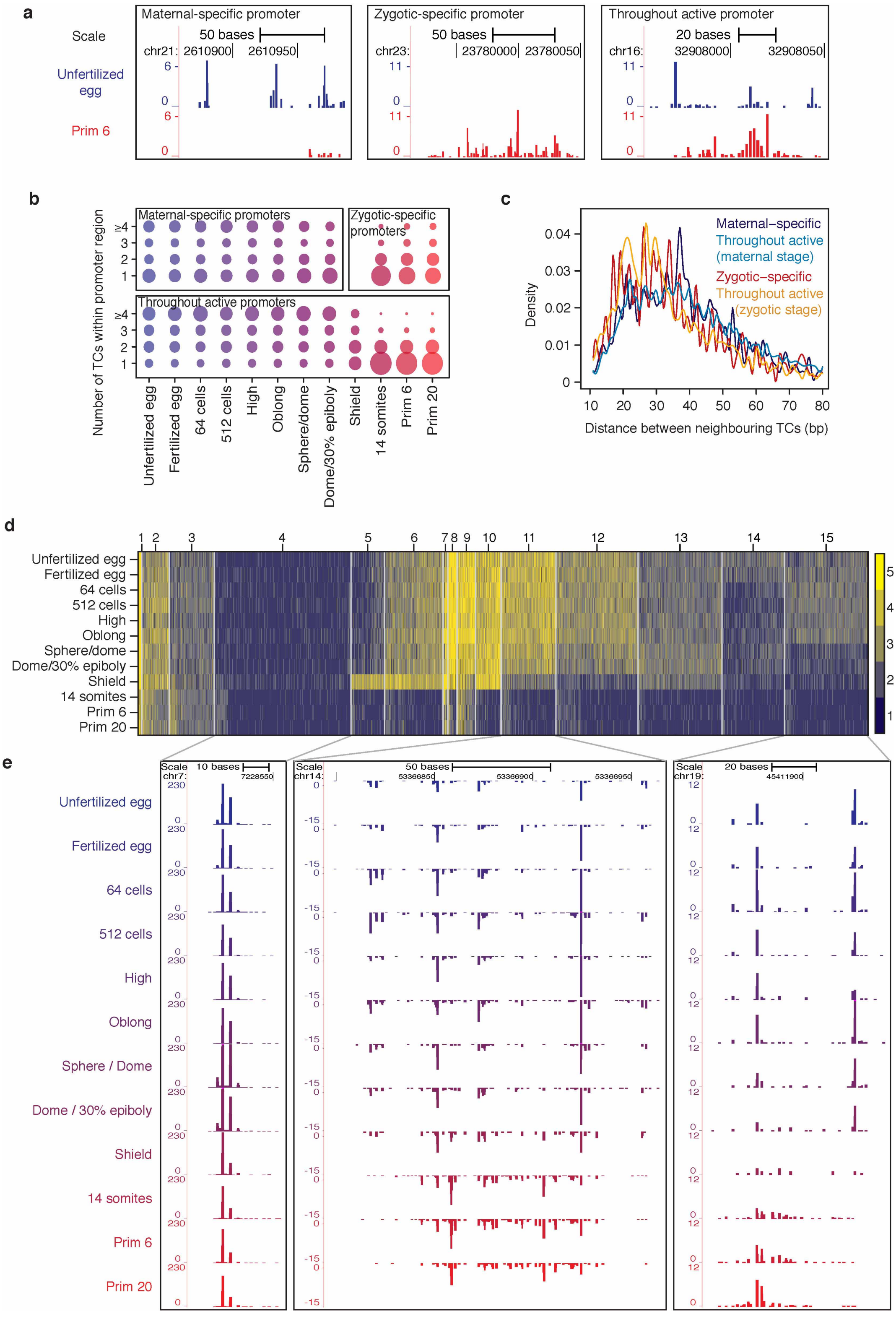
Extended Data Figure 2 | Definition and properties of shifting promoters. **a**, Schematic representation of shifting promoter score calculation between two stages. Forward and reverse scores are calculated using forward (F1, F2) and reverse (F1′, F2′) cumulative sums of CAGE tags along the promoter region, and the final shifting score is the maximum of the two. **b**, Heatmap showing the number of shifting promoters detected by applying the method shown in **a** (using a 0.6 score threshold) to all possible pairs of developmental stages. The majority of shifting was detected between the first six stages (predominantly maternal) and last three stages (predominantly zygotic). Promoters that qualified as shifting in at least four of these early versus late pairs were selected for a final set of 911 shifting promoters. **c**, Distribution of distance between maternal and zygotic dominant TSS (shift) for 911 shifting promoters. **d**, Developmental expression dynamics of maternal (blue) and zygotic (red) dominant TSS in shifting promoters. Average expression for the entire promoter is shown in black. **e**, Density of AA, AT, TT and CG dinucleotides for shifting promoters sorted by magnitude of shift. Promoters were centred at either the maternal (left) or the zygotic (right) dominant TSS. Blue arrowhead marks the AA/AT/TT enrichment ~30 bp upstream of the maternal TSS and red arrowhead marks the boundary between CG and AA/AT/TT enrichment ~50 bp downstream of the zygotic TSS. **f**, Relative frequency of top ten most frequent TATA-like pentamers in the region −35 to −22 bp upstream of maternal (blue) and zygotic (red) dominant TSS. Percentage of match to TATA-box PWM is denoted in the brackets.



Extended Data Figure 3 | Genome-wide sequence and chromatin signature of maternal versus zygotic TSSs. **a**, Dinucleotide density for all throughout-active promoters sorted by orientation-sensitive distance between the maternal and zygotic dominant TSS. Promoters were centred at either the maternal (left) or the zygotic (right) dominant TSS. **b**, Distribution of match (%) to TATA-box PWM in the region −35 to −22 bp upstream of the maternal (blue) and zygotic (red) dominant TSS in constitutively active promoters. *P* value of the two-tailed Wilcoxon rank-sum test is shown. **c**, Relative frequency of top ten most frequent TATA-like pentamers upstream of the maternal (blue) and zygotic (red) dominant TSS. Percentage of match to TATA-box PWM is denoted in the brackets. **d**, H2A.Z signal in the prim 6 stage at the same set of promoters as in **a**. Subtracted coverage of reads mapping to the plus and minus strand is shown. **e**, Density of TT dinucleotide in a ±100 bp region for the same set of promoters as in **a**. **f**, WW dinucleotide density for maternal- (left) and zygotic-specific (right) promoters centred at their dominant TSS and sorted by decreasing match to the TATA-box (white dashed lines). Average density of individual dinucleotides is shown below. **g**, Distribution of match (%) to TATA-box PWM in the region −35 to −22 bp upstream of dominant TSS for maternal- (blue) and zygotic-specific (red) promoters. *P* value of the two-tailed Wilcoxon rank-sum test is shown. **h**, Relative frequency of the top ten most abundant TATA-like pentamers in maternal (blue) and zygotic (red) promoters. Percentage of match to TATA-box PWM is denoted in the brackets.

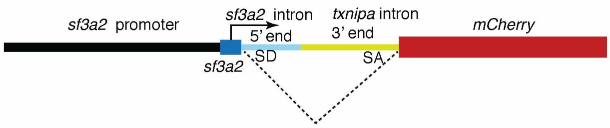


Extended Data Figure 4 | Composite multiple sharp architecture of maternal-specific promoters. **a**, Distribution of promoter interquantile width ($q_{0.1} - q_{0.9}$; central promoter region that contains ≥80% of CAGE tags) for 1,944 maternal-specific promoters. **b**, Number of TCs within the promoter region for maternal broad promoters (interquantile width >20 bp) across early developmental stages. Bubble size reflects relative number of promoters that contain given number of TCs. **c**, Comparison of dinucleotide frequencies around single dominant TSS in promoter (left) and dominant TSSs from all TCs within the same promoter (right). **d**, Composite architecture of a maternal broad promoter showing multiple sharp TCs with their associated W-box at the expected ~30 bp upstream position.



Extended Data Figure 5 | Change of promoter architecture during maternal to zygotic transition. **a**, Examples of maternal-specific, zygotic-specific and throughout-active promoters showing difference in architecture between maternal and zygotic stage. **b**, Number of TCs within promoter region for maternal-specific, zygotic-specific and constitutively active promoters. Bubble size reflects relative number of promoters that contain given number of TCs and different colours correspond to different developmental stages. **c**, Distribution of distance between neighbouring TCs within the same promoter region. **d**, Heatmap showing *k*-means clustering ($k = 15$) of throughout-active promoters, based on the number of TCs that they contain throughout 12 developmental stages. **e**, Representative examples for three selected clusters: highly expressed sharp promoter whose architecture does not change and retains only one TC (left); promoter in which number of TCs decreases from five clearly separated sharp TCs in maternal stages to one fairly broad TC in zygotic stages (middle); promoter that changes its architecture from two clearly separated sharp TCs to one broader TC after maternal to zygotic transition (right).

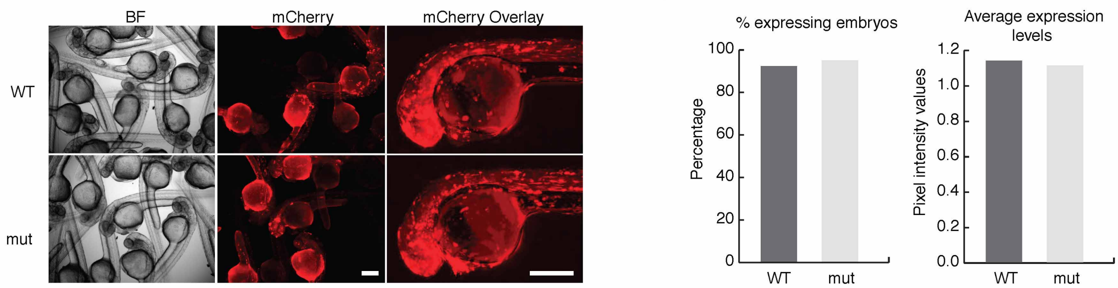
a



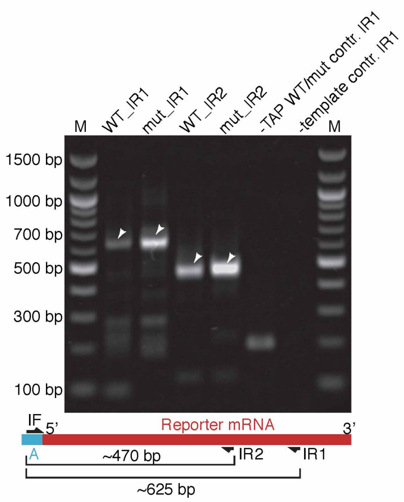
b

Primer		Sequence
sf3a2 promoter region wt	F	aggatccatactcgaTCTCCAACCCTCACATGTAAAAAC
	R	tcaacataGCAGACACAGATCAACAACAGC
txnipa intron	F	gtgtctgcTATGTTGATGAGATTGGATGTGG
	R	ctcaccatggctcgaCTGTGAAAAATAAAGATACATACATTA
sf3a2 promoter region site-directed mutagenesis	F	TTTATGCTGAcTTgATCAcTAcAGTTGTGCGTGTCAAcATcACAGGAAgAGGAAGTGCATCATgTGGCGCATGCG
	R	AAGTCAGCATAAAAGTGTATTC

c **d**



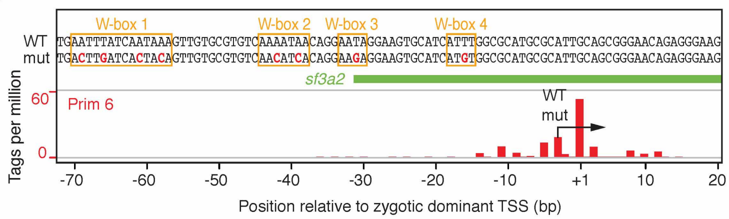
e



f

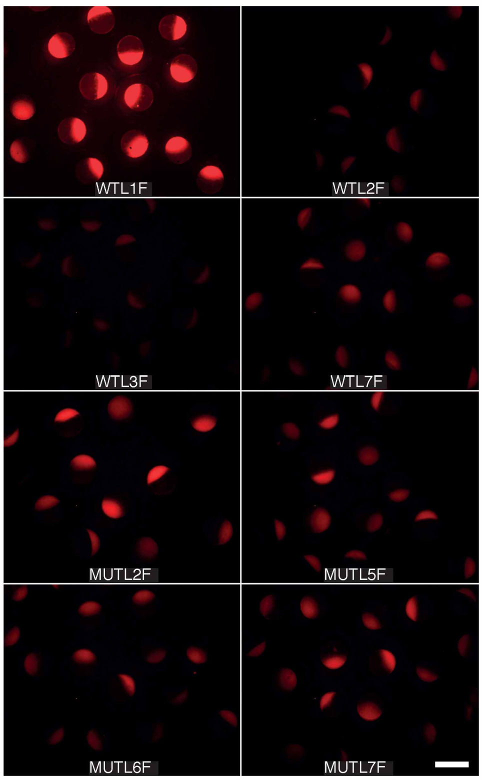
Primer	Sequence
Outer PCR forward primer	GCTGATGGCGATGAATGAACACTG
Inner PCR forward primer	CGCGGATCCGAACACTGCGTTTGCTGGCTTTGATG
Outer PCR *mCherry* reverse primer	TTACTTGTACAGCTCGTCC
Inner PCR *mCherry* reverse primer 1 (IR1)	CTTCAGCTTCAGCCTCTGCT
Inner PCR *mCherry* reverse primer 2 (IR2)	CTTCACCTTGTAGATGAA

g

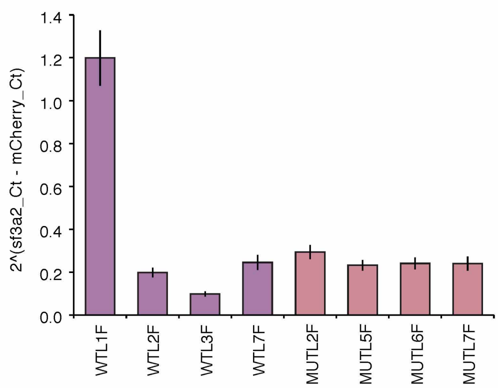


Extended Data Figure 6 | Validation of *sf3a2* promoter code in transgene injected embryos. **a**, Schematic of reporter construct; the dashed lines indicate the spliced out intronic sequence. SA, splice acceptor; SD, splice donor. **b**, Primers used for *sf3a2* construct. **c**, Reporter gene expression in the wild-type and mutant *sf3a2* construct-injected embryos (36 hpf). Bright field (BF) and red channel fluorescent images (mCherry) are shown. Scale bar, 250 μm. The panels on the right show red channel projection overlay of 24 (wild type (WT), top) and 31 (mutant (mut), bottom) embryos. **d**, Percentage of mCherry-expressing embryos from total ECFP-positive embryos (left) and overall reporter activity measured as average pixel intensity values from 24 wild-type and 31 mutant *sf3a2* promoter-injected embryos (right). **e**, 5′ RACE of *sf3a2* promoter construct. Fragments amplified by PCR using a forward primer specific for the 5′ RACE adaptor and two different inner reverse primers specific for the mCherry reporter. The arrowheads indicate the PCR products of expected size. A, 5′ RACE adaptor; IF, inner forward primer; IR1, inner reverse primer 1; IR2, inner reverse primer 2; M, DNA marker; mut, embryos injected with construct containing *sf3a2* promoter with mutations disrupting W-boxes (as shown in **f**); TAP, tobacco acid pyrophosphatase; WT, embryos injected with construct containing wild-type *sf3a2* promoter. **f**, Primers used in 5′ RACE. **g**, Genomic sequence of the *sf3a2* promoter with point mutations introduced to disrupt W-boxes (orange) marked in red. The track at the bottom shows TSSs detected by CAGE in wild-type zebrafish embryos in prim 6 zygotic stage. Black arrow marks TSS confirmed by 5′ RACE in the zygotic stage in both the wild-type and the mutated *sf3a2* promoter.

a



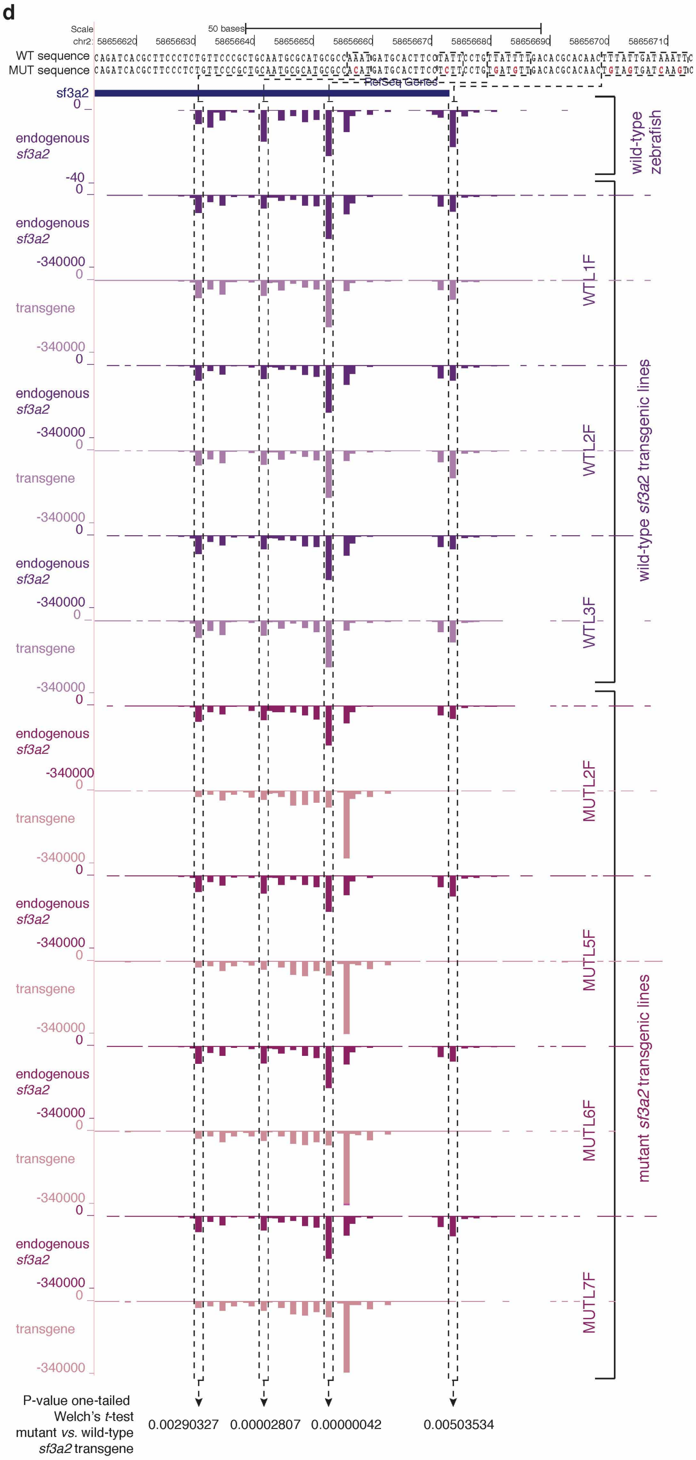
b



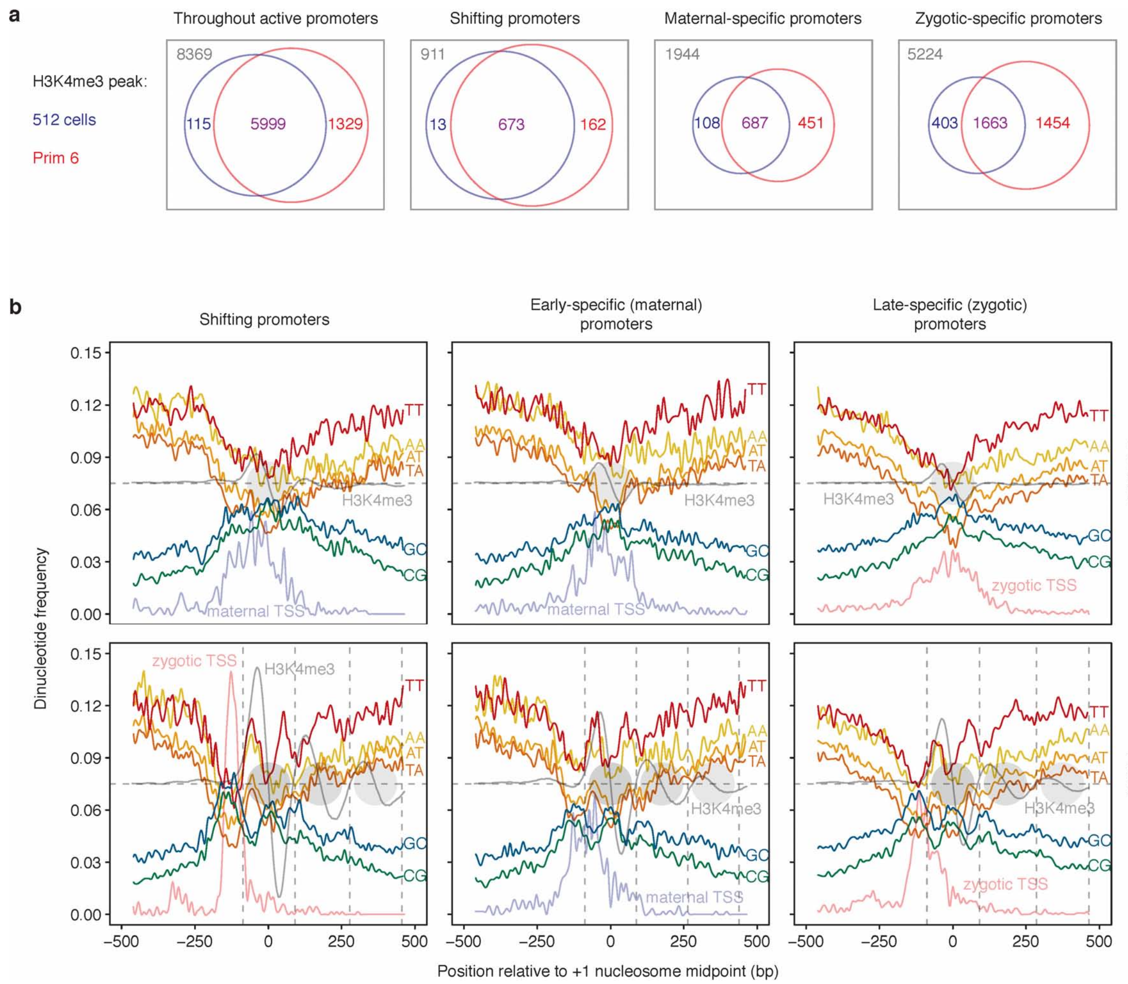
c

		mChery transgene	endogenous *sf3a2*
pair 1	F	GGAACAGAGGGAAGCGTGATCT	GCGTTAGAGACCATCGACATCAA
	R	CACCTTGAAGCGCATGAACTCC	CCCCTCATTATTGTGCAGCGTG
pair 2	F	CAGAGGGAAGCGTGATCTGGAC	GCGTTAGAGACCATCGACATCAATAA
	R	CTCCATGTGCACCTTGAAGCG	AGTGTGTGCAAGATAACTCCCCT

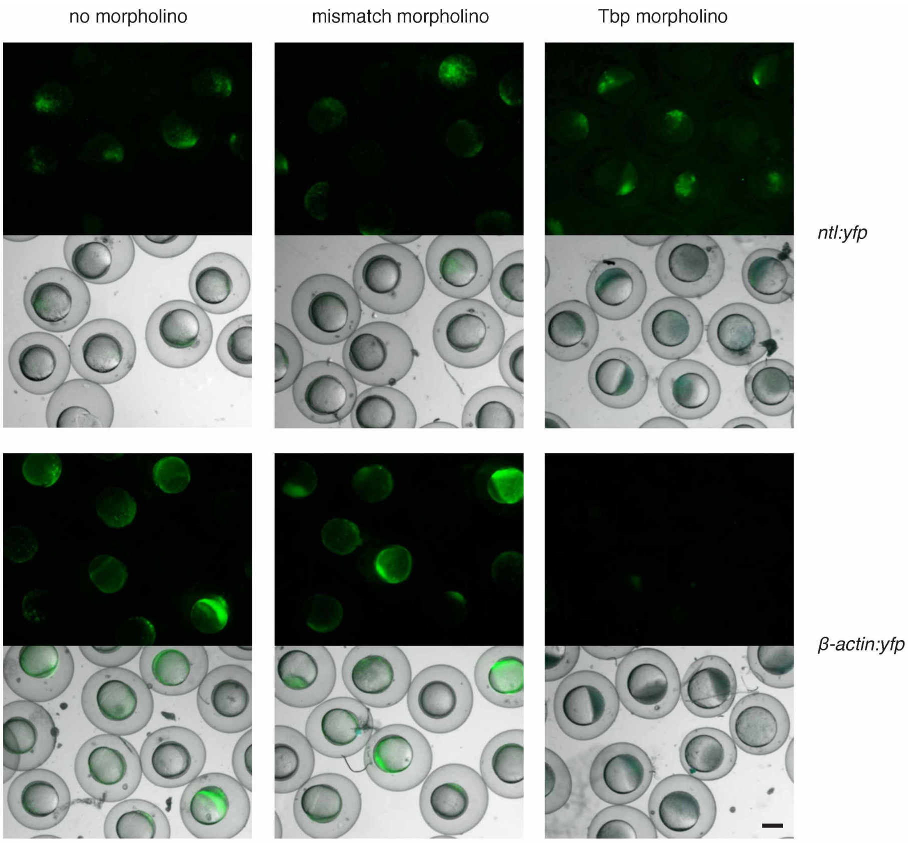
d



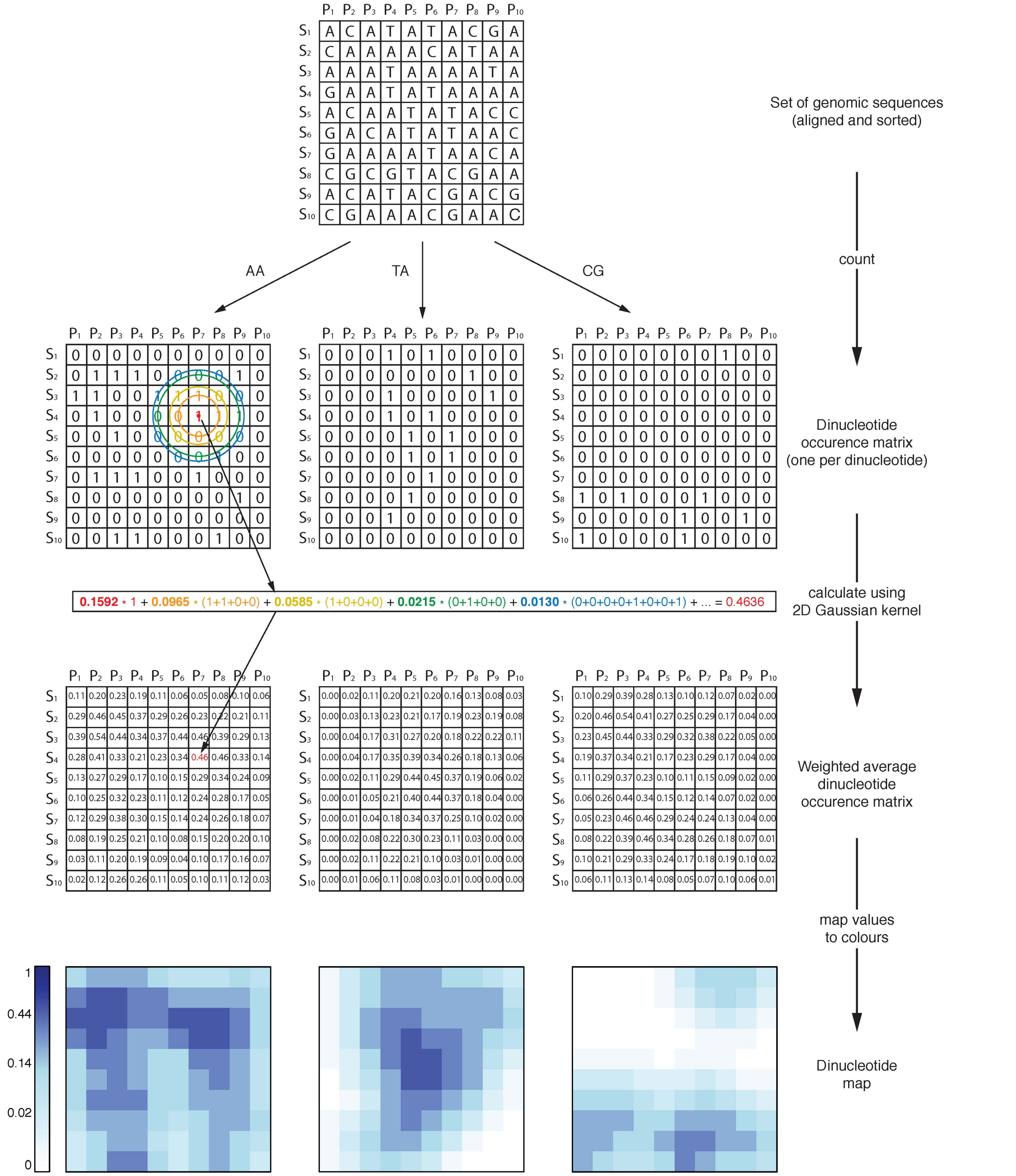
Extended Data Figure 7 | Validation of *sf3a2* promoter code in stable transgenic lines. **a**, *sf3a2:mCherry* transgene expression in wild-type (WT; top four panels) and mutant (MUT; bottom four panels) *sf3a2* promoter transgenic embryos (high stage). Scale bar, 1 mm. **b**, Quantitative polymerase chain reaction (qPCR) results of transgene expression: ratio between abundance of *sf3a2:mCherry* trasgene mRNA and endogenous *sf3a2* mRNA. Bar height shows mean value across six qPCR experiments (three replicas for each of the two different primer pairs) and error bars denote 95% confidence interval of the mean. **c**, Sequence of primers used in qPCR. **d**, TSSs and their relative usage detected by SL-CAGE in three wild-type and four mutant *sf3a2* promoter transgenic lines. Promoter sequence is shown at the top, with point mutations introduced to disrupt W-boxes highlighted in red. For each transgenic line two tracks are shown: TSSs used by the endogenous *sf3a2* promoter (upper track, more intensive colour) and TSSs used by the transgene (lower track, lighter colour). *P* value (one-tailed Welch's two sample *t*-test) of the difference between mutated and wild-type promoter variants is shown for four maternal TSSs associated with disrupted W-boxes (dashed boxes).



Extended Data Figure 8 | H3K4me3-marked nucleosome positioning signal in maternal and zygotic stage. **a**, Venn diagrams showing number of promoters containing H3K4me3 peak within ±1 kb of TSS in early (512 cells) and late (prim 6) stage for distinct classes of promoters (as defined in Extended Data Fig. 1a). **b**, Frequency of selected dinucleotides centred at the position of the first downstream (+1) nucleosome for distinct classes of promoters. Only promoters containing an H3K4me3 peak in both 512 cells and prim 6 stages were used. Centres of nucleosomes were estimated from subtracted H3K4me3 coverage (shown in grey). Density of TSSs in maternal (512 cells) and zygotic (prim 6) stages is shown in light blue and light red, respectively. Note that maternal-specific promoters have only maternal TSSs and the distribution of these TSSs in the zygotic stage is shown only for orientation. The opposite is true for zygotic-specific promoters.



Extended Data Figure 9 | Validation of Tbp morphant phenotypes before ChIP-seq analysis. Tbp morphants show epiboly defects and differential loss of reporter activity as previously described[42,43]. Embryos in the 30% epiboly stage showing epiboly delay and corresponding Tbp-independent (*ntl:yfp*) and Tbp-dependent (*β-actin:yfp*) reporter gene activities were used for H3K4me3 ChIP-seq analysis. Zebrafish fertilized eggs were injected with the reporter constructs as indicated on the right followed by injection of morpholino (Tbp 5 mismatch or Tbp mo1 morpholino[42], respectively). Fluorescence images on top, bright field images below. *ntl*, notail; *yfp*, yellow fluorescent protein. Scale bar, 400 μm.



Extended Data Figure 10 | Dinucleotide maps: calculation and visualization of dinucleotide patterns. Genomic sequences (of the same length) are sorted and aligned into a matrix-like representation (P_1–P_{10}, positions; S_1–S_{10}, sequences). Marking the presence of selected dinucleotide (for example, AA) by 1 and the absence by 0 creates an occurrence matrix. Next, a weighted average is calculated at each position by placing a two-dimensional Gaussian kernel at that position and assigning weights to surrounding positions. An example of calculating the value at position S_4,P_7 is shown. Surrounding positions are coloured on the basis of the weights assigned to them by the Gaussian kernel (bandwidth = 1 in both dimensions, and covariance = 0 between the two dimensions). Averaged values are mapped to different shades of blue to visualize the dinucleotide density across the set of input sequences.

CORRECTIONS & AMENDMENTS

ADDENDUM

doi:10.1038/nature13077

Addendum: Optical-field-induced current in dielectrics

Agustin Schiffrin, Tim Paasch-Colberg, Nicholas Karpowicz, Vadym Apalkov, Daniel Gerster, Sascha Mühlbrandt, Michael Korbman, Joachim Reichert, Martin Schultze, Simon Holzner, Johannes V. Barth, Reinhard Kienberger, Ralph Ernstorfer, Vladislav S. Yakovlev, Mark I. Stockman & Ferenc Krausz

Nature **493**, 70–74 (2013); doi:10.1038/nature11567

Calculations in this Letter[1] and in ref. 2 are sensitive to the dipole matrix elements, $Z_{vi,cj}$ $(i, j = 1, 2)$, describing the transitions between the two uppermost valence ($v1$ and $v2$) and two lowest conduction ($c1$ and $c2$) bands (see Supplementary Information to ref. 1). They have resulted in good agreement with all observations of both experiments for $Z_{v1,c1} = 3$ Å, $Z_{v1,v2} = Z_{c1,c2} = 2$ Å and $Z_{v2,c2} = 0$. By mistake, Fig. 1b and Supplementary Fig. 4 of ref. 1 show adiabatic levels of silica for $Z_{v1,c1} = 1$ Å ($Z_{v1,v2}$, $Z_{c1,c2}$ and $Z_{v2,c2}$ are correct and set as above). Here we correct this error in Fig. 1 by plotting the adiabatic levels of silica for the value of $Z_{v1,c1} = 3$ Å that we actually used. This corrected level diagram prompts a correction of the qualitative picture mediated by Fig. 1b and Supplementary Figs 4 and 6 of ref. 1, aiming at a qualitative insight of the strong-field optical phenomena reported in refs 1 and 2. The corrected picture suggests an analogy to a phenomenon known in semiconductor physics for more than two decades. We summarize these findings below. The experimental and quantitative results presented in refs 1 and 2 stand as published, as do the conclusions derived directly from them.

As we have argued in the Supplementary Information to ref. 1 for slowly varying fields of strengths $|F| \approx 1-3$ V Å^{-1}, electronic states are predicted to undergo Wannier–Stark localization[3] within approximately one unit cell of the wide-gap material—silica—under scrutiny. At a field strength where a level anticrossing denoted by Δl occurs in Fig. 1, Zener tunnelling[4] between the localized Wannier–Stark states across $|\Delta l| - 1$

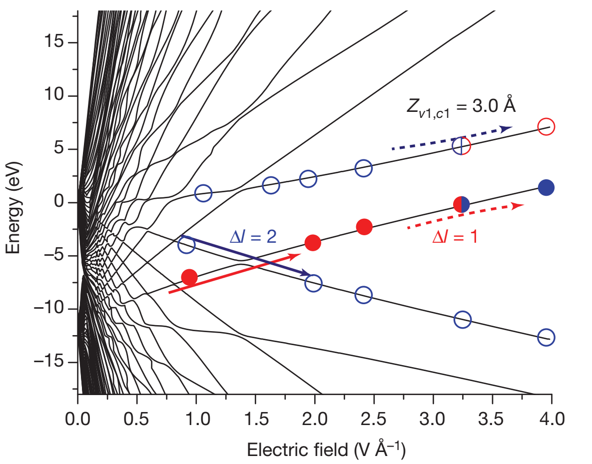


Figure 1 | Electronic transitions between the valence and conduction bands of silica in a strong optical field. Schematic of the adiabatic energy levels as a function of external electric field, for interband dipole moment $Z_{v1,c1} = 3$ Å. A crystal with a thickness of 50 unit cells in the field direction is considered. The anticrossings with $\Delta l = 1$ and $\Delta l = 2$ are indicated. Red circles denote levels with the valence-band wavefunctions; blue circles denote levels with the conduction-band wavefunctions. Closed circles indicate filled states and open circles indicate empty states. The arrows indicate level anticrossings passed diabatically (solid crossed arrows) and adiabatically (dashed parallel arrows).

unit cells is expected to occur, with a rate equal to the anticrossing energy gap (in frequency units) $\Delta E_l/(2\pi\hbar)$. As Fig. 1 reveals, for $|\Delta l| \geq 2$, this rate is very small with respect to the carrier frequency, $\omega_L/2\pi$, of the optical field (where subscript 'L' is for 'laser'). Hence, all these anticrossings are passed diabatically in the oscillating strong laser field. The system follows the diabatic terms (shown by the solid crossed arrows in Fig. 1) as the field increases or decreases without significant transfer of population between the valence band and the conduction band. This is because the system passes the anticrossings too quickly for Zener tunnelling to occur. Such an anticrossing is mostly inconsequential (almost 'ignored' by the system).

As the corrected level diagram in Fig. 1 reveals, the gap becomes very large, with $\Delta E_l/\hbar >> \omega_L$, for the ultimate level anticrossing at $|\Delta l| = 1$ occurring when the field amplitude approaches and exceeds the critical field strength $F_{crit} = \Delta_g/ea$ where $\Delta_g \approx 9$ eV is the bandgap of silica, e is the charge on the electron and a is the lattice constant). The large anticrossing gap implies a large probability of Zener tunnelling between the valence band and conduction band states, which are Wannier–Stark-localized at the nearest neighbours. This anticrossing is passed predominantly adiabatically, as indicated by the dashed parallel arrows in Fig. 1. This leads necessarily to the lower anticrossing level gradually changing its wavefunction from valence band to conduction band (the well known exchange of quantum numbers in adiabatic anticrossings) while remaining fully occupied. This gradual change can be quantified by projecting the actual filled, adiabatic field-modified valence band states to zero-field conduction-band states as given by Supplementary equation (10) in ref. 2. This projection results in a transient population of virtual conduction-band electrons (the evolution of which is shown in figure 3d of ref. 2) and the corresponding virtual valence-band holes.

These are virtual carriers similar to those responsible for virtual photoconductivity in semiconductors theoretically predicted[5–7] and experimentally observed[8] over two decades ago. However, the analogy is incomplete. In our case, the virtual carriers are produced by a non-resonant (mostly adiabatic), strong-field (and hence, necessarily, non-perturbative) excitation, in contrast with early work[5–8], where they were excited via a near-resonant perturbative excitation. These virtual carriers make the system more polarizable in a static electric field[5–8] and likewise at optical frequencies. If so, they can account for both the transient reflectivity shown in figure 3b of ref. 2 and the transient current displayed in figure 3 of ref. 1. In both cases, a strong few-cycle field nonlinearly enhances the strong-field polarization of the dielectric by creating virtual carriers. These respond with increased excursions to a weaker optical field, which probes the resultant enhancement in reflectivity and separates the charges that—after being trapped in the electrodes or within the insulator—subsequently induce a current through an external circuit in refs 2 and 1, respectively. Note that our quantitative quantum model[1,2] fully takes into account this virtual-carrier mechanism.

We emphasize that our original interpretation[1,2] of the strong field causing a reversible quantum transition (that we name 'semi-metallization') from an insulating state to one with increased conductivity is valid. It allows one to gain a useful independent insight into highly non-equilibrium non-perturbative processes in strong ultrafast optical fields (see also ref. 9). We note that recent *ab initio* calculations[10] for insulators in strong fields are interpreted in terms of the field-induced insulator to metal transition.

In conclusion, we have provided a unified qualitative explanation for the optical-field-induced transient reflectivity[2] and current[1] in terms of virtual electron–hole pairs generated by a strong optical field. In sharp contrast with virtual photoconductivity of semiconductors[5–8], these effects are induced by photons far below the bandgap and respond nearly adiabatically—in a highly nonlinear, nonperturbative fashion—to the instantaneous field, rather than its cycle-averaged intensity as in refs 5–8, as it approaches F_{crit}. This opens up the prospect of exploiting the new effects for advancing solid-state high-speed metrology and signal processing towards the petahertz frontier.

1. Schiffrin, A. *et al.* Optical-field-induced current in dielectrics. *Nature* **493,** 70–74 (2013).
2. Schultze, M. *et al.* Controlling dielectrics with the electric field of light. *Nature* **493,** 75–78 (2013).
3. Wannier, G. H. *Elements of Solid State Theory* (Cambridge Univ. Press, 1959).
4. Zener, C. A theory of the electrical breakdown of solid dielectrics. *Proc. R. Soc. Lond. Ser. A* **145,** 523–529 (1934).
5. Yamanishi, M. Field-induced optical nonlinearity due to virtual transitions in semiconductor quantum well structures. *Phys. Rev. Lett.* **59,** 1014–1017 (1987).
6. Chemla, D. S., Miller, D. A. B. & Schmitt-Rink, S. Generation of ultrashort electrical pulses through screening by virtual populations in biased quantum wells. *Phys. Rev. Lett.* **59,** 1018–1021 (1987).
7. Yablonovitch, E., Heritage, J. P., Aspnes, D. E. & Yafet, Y. Virtual photoconductivity. *Phys. Rev. Lett.* **63,** 976–979 (1989).
8. Hu, B. B., Zhang, X. C. & Auston, D. H. Terahertz radiation induced by subband-gap femtosecond optical excitation of GaAs. *Phys. Rev. Lett.* **67,** 2709–2712 (1991).
9. Krausz, F. & Stockman, M. I. Attosecond metrology: from capturing electrons toward speeding up signal processing. *Nature Photon* **8,** 205–213 (2014).
10. Wachter, G. *et al.* Ab-initio simulation of optical-field induced currents in dielectrics. Preprint at http://arxiv.org/abs/1401.4357 (2014).

Correspondence should be addressed to A.S. (agustin.schiffrin@mpq.mpg.de), M.I.S. (mstockman@gsu.edu) or F.K. (ferenc.krausz@mpq.mpg.de).

ERRATUM

doi:10.1038/nature13142

Erratum: Global carbon dioxide emissions from inland waters

Peter A. Raymond, Jens Hartmann, Ronny Lauerwald, Sebastian Sobek, Cory McDonald, Mark Hoover, David Butman, Robert Striegl, Emilio Mayorga, Christoph Humborg, Pirkko Kortelainen, Hans Dürr, Michel Meybeck, Philippe Ciais & Peter Guth

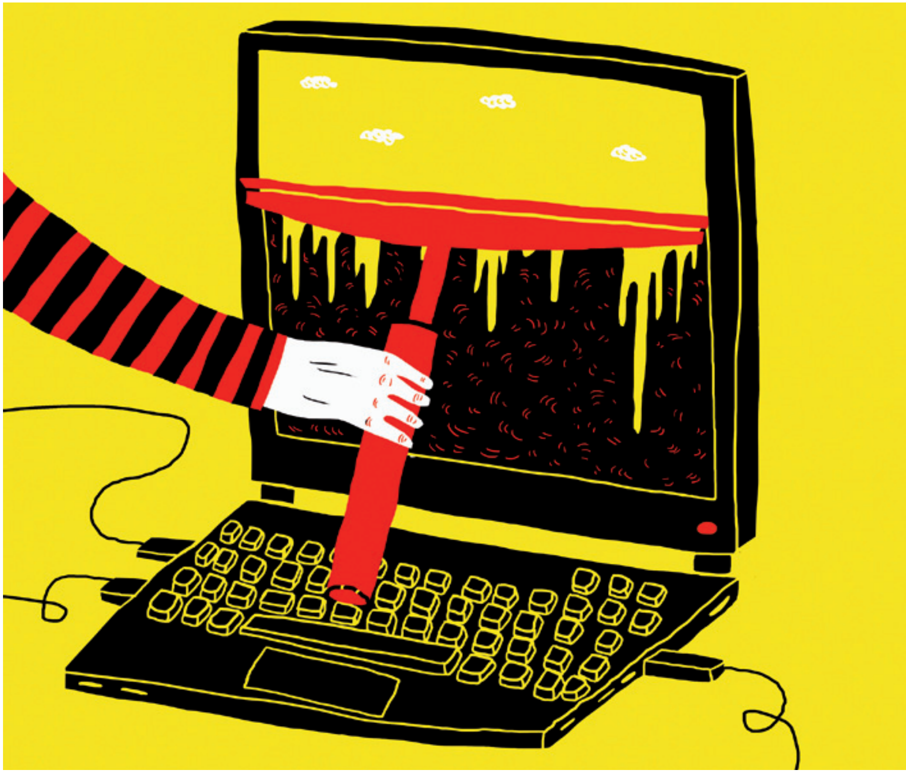
Nature **503,** 355–359 (2013); doi:10.1038/nature12760

In Fig. 1a of this Article, the highest value on the colour scale should be 11,772 instead of 1,772. This has been corrected in the HTML and PDF versions of the paper online.

COLUMN Helping graduate students to face their myriad financial challenges **p.391**

FACEBOOK The hottest science-careers information **https://www.facebook.com/naturejobs**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**



OIVIND HOVLAND/IKON IMAGES/GETTY

RETRACTIONS

A clean slate

Mistakes are part of science. But setting the record straight promptly and clearly can help to avoid a career blot.

BY VIRGINIA GEWIN

After 18 months of complex testing and retesting, Pamela Ronald became certain that she needed to retract two high-profile papers on disease resistance in rice[1,2]. The hardest part, says Ronald, a crop scientist at the University of California, Davis, was staying calm — she worried about the implications for current and past lab members and about others spending time replicating potentially faulty work.

The papers had claimed to identify a bacterial protein that could activate an immune response in rice plants with a specific receptor. But when new members of her team were unable to reproduce the results, alarm bells started ringing. Shaken, they decided that the first step was to genotype all the laboratory strains in their collection. Eventually they caught a labelling error: two of the 12 strains thought to lack the protein in question actually lacked a different protein. And the careful backtracking unearthed yet another error: the test, which they had used to verify that this protein could trigger resistance, turned out to be faulty. Despite her distress, throughout the ordeal Ronald was straightforward with journal editors and her colleagues about the likelihood of retractions. She knew that her scientific reputation depended on complete transparency about possible errors. "You just have to set aside emotions and let the scientific process pull you through," she says.

Worldwide, retractions are on the rise: last year alone, scientific journals retracted roughly 500 papers (of more than 1 million published), compared with fewer than 50 per year in the early 2000s (see *Nature* **478,** 26–28; 2011).

One study[3] — in the life sciences — suggests that misconduct, such as plagiarism or falsified data, has been to blame for two-thirds of retractions (see *Nature* **490,** 21; 2012). And behavioural ecologist Daniele Fanelli of the University of Montreal in Canada, who studies the issue, says that at least one-quarter are the result of unfortunate mistakes. The rise in retractions could be because scientists are making more errors, but it could also indicate a growing culture of coming clean on errors. And that, Fanelli says, is a positive trend. "We really need to think more about how to reward retractions that are correcting mistakes — find a way to make them a badge of honour instead of a badge of shame," he says.

Scientists often treat retractions as dirty secrets. The muted discourse means that the process is often much more confusing, frustrating and embarrassing for researchers, journal editors and universities than it needs to be. Many struggle with the best way to correct the record and with how to salvage viable data. Yet if a retraction is the result of an accident or honest error, it should not be a blot on an otherwise respectable publication record. Scientists and journal editors who have retracted papers say that the process can be handled productively, whether the errors are from contamination, a cell-line mix-up or statistical analyses gone awry. Above all, they say, transparency is key.

SENSING A PROBLEM

A decade ago, retractions were far from transparent. "It was not unusual to see 'Paper is retracted' [in the journal] and nothing else," says Ferric Fang, a microbiologist at the University of Washington in Seattle who also studies retractions. In 2009, realizing that journals lacked policies, the UK-based non-profit organization COPE (Committee on Publication Ethics) published guidelines on how best to correct the scientific record. COPE expects its more than 9,000 journal members to follow the guidelines, which recommend that retraction statements link to the retracted article, be freely available and state who is retracting the article, among other criteria. But perhaps most importantly, says former COPE chair Liz Wager, a retraction notice should include the reason that the retraction was made to clearly distinguish misconduct from honest error. "The retraction has to be complete and ▶

honest, and clearly articulate what things are wrong," says Inder Verma, a molecular biologist at the Salk Institute for Biological Studies in La Jolla, California, and editor-in-chief of *Proceedings of the National Academy of Sciences*.

RETRACTION BREAKDOWN

There are several ways to amend a published paper (see 'Retraction guide'). A retraction is reserved for the most severe problems, ones that unravel a paper's conclusions. It is hard to retract a paper and later republish the valid parts, says Arturo Casadevall, editor-in-chief of *mBio* and a microbiologist at the Albert Einstein College of Medicine in Bronx, New York. "Many journals won't allow that."

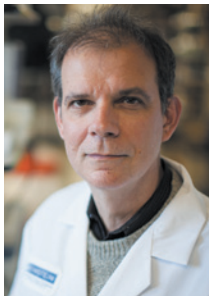
For less serious issues, there are two other options. For the least egregious errors — a mislabelled figure, for example — a correction often suffices. Alternatively, the paper can be partially retracted if the erroneous findings do not invalidate the article's stated conclusions. (COPE discourages partial retractions, however, saying that they "make it difficult for readers to determine the status of the article and which parts may be relied upon".)

Once alerted to a potential problem, the first step is to identify whether an error has actually occurred and, if so, how. The authors and journal editors then need to decide on the appropriate response. As soon as Ronald realized something was amiss with her original work, she contacted the editors of the journals that had published the papers. "Part of you wants to retract it immediately," she says. But she knew that she needed to ascertain the facts first.

Her team worked out all the experiments that would be needed to determine the problem. "Everyone had slightly different concerns, but we all wanted to get it right," she says. Finding the error became all-consuming; other projects languished.

And Ronald went a step further. Once she was sure that there was a problem, she contacted colleagues to highlight the issue and gave a public seminar to inform the international community. "I was alarmed that others were trying to build on this work when we couldn't, and I didn't want to waste anyone's time," she says. Her efforts won her praise both from her colleagues and from the *Retraction Watch* blog, which reports on scientific retractions and misconduct cases. In a post about Ronald's case, *Retraction Watch* co-founder Ivan Oransky said that the blog likes "being able to point out when researchers stand up and do the right thing, even at personal cost".

Deciding whether to issue a full-blown retraction is difficult. Reputations are on the line, and authors may not necessarily agree on how best to amend the mistake. It can take a lot of negotiating with journal editors and colleagues.



"The greatest currency we have as scientists is respect."
Arturo Casadevall

That was the experience of Daniel St Johnston, a developmental biologist and director of the Gurdon Institute at the University of Cambridge, UK. He and his colleagues had published two papers[4,5] describing a pathway that they thought was essential for the epithelial cells surrounding a developing egg to maintain their polarity, or orientation, under starvation conditions. However, they later discovered that they had damaged some of the cells when they had dissected the tiny ovaries of the starved fruit flies, resulting in 'false clones' that mimicked the appearance of the cells of interest, but that did not have the same mutations.

St Johnston first wrote a new paper highlighting the existence of false clones. Once that paper was accepted in *Biology Open*[6], he sent the manuscript to the two journals that had published the original incorrect papers, *Journal of Cell Biology* and *Developmental Cell*, in a bid to find the best way to link the new findings to the original papers. The journal editors advised him to retract the original papers because the main conclusions were no longer valid. Then things got complicated.

One author was not happy about the prospect of losing the data that remained sound from the scientific record. But to retract the papers, all authors had to approve. The researchers and journal editors discussed various options, including a partial retraction, but couldn't reach an agreement. To break the deadlock, St Johnston arranged for the valid data to be published alongside the false-clone findings in *Biology Open*[7], and the retractions went ahead. St Johnston says that he had to figure out the best course as he went along: "There's no obvious mechanism in place to handle the useful, valid content of a retracted paper." He adds that, so far, he has not experienced any career backlash.

SURMOUNTING STIGMA

The data support what many journal editors advise — that the best way to overcome the stigma associated with a retraction is to come clean with a detailed account of mistakes. A study published last November[8] found that authors who self-report mistakes and retract papers accordingly will not lose out on citations to their previous, legitimate research. However, authors who fail to self-report the need for a retraction lose up to 12.5% of citations per year per paper, five years after the retraction, compared with non-retracted papers with similar citation patterns. "Being transparent preserves your reputation as someone honestly seeking truth," says study author Benjamin Jones, a higher-education researcher at Northwestern University in Evanston, Illinois.

In fact, a clear retraction may end up being worthy of a citation itself. For example, a 2006 retraction[9] detailed an author dispute over whether original findings of ferromagnetic behaviour found in a carbon-60 atom remained valid in the face of new measures. "I've found examples of retracted papers that continue to be cited because the retraction is clear, suggesting that people in the field are more discerning than they've been given credit for," says Fang.

In St Johnston's experience, the more common case is that scientists bury the error — which, he says, harms the scientific community by squandering everyone's time and resources. "I personally didn't want to do that because my reputation is my trademark," he says. "If I own up to my mistakes, people know that if I get it wrong, I will tell them."

St Johnston knew that retracting the paper was the right thing to do, but he says that if he had been at an earlier, more precarious stage of his career, he might have thought twice. And he did have serious concerns about the impact it would have on his junior co-authors. "Most of the scientific community doesn't bother

RETRACTION GUIDE

Publishers have a range of options for amending the scientific record, depending on the circumstances.

Action	Example	Paper conclusions affected	Issued by
Correction (erratum)	Typographical errors Mislabelled figures Author/contributor list incorrect	No	Paper author (ideally)
Expression of concern	Investigation suggests that data are not reliable Suspicions of misconduct	Unsure	Editor
Partial retraction	Figure or table based on corrupt data Some data inappropriately analysed	Yes, but not overall finding	Author or editor
Retraction	Clear evidence of misconduct Error invalidates work	Yes	Author or editor

SOURCE: COPE; COUNCIL OF SCIENCE EDITORS

distinguishing between an error and something more sinister," he says.

Indeed, some data suggest that early-career researchers are right to be concerned. Jones has compared how the citation records of eminent and less-well-established authors are affected after a retraction[10]. "If you are highly established, we see little effect if you are one of the authors on a retracted paper. If you are not well established, you will see a negative effect," he says.

As a then-assistant professor, Geoffrey Chang had such concerns when he was alerted to a problem with his work, and threw himself into discovering the error. A crystallographer at the University of California, San Diego, he had had a string of successes in the form of awards, grants and high-profile papers starting in 2001, when he was in his late 20s.

When colleagues called his findings into question, Chang pored over his home-made computer program and eventually discovered that he had transposed two data columns. He retracted five papers, and republished the corrected molecular structures in two papers roughly a year after pinpointing the problem. The thorough, transparent reanalysis of the data, he says, corrected the scientific record, which helped to maintain his academic standing. "It was seven years ago," he says, "and it's still very much a part of me. But it made us a stronger, more careful lab." Chang has since secured major grants, including grants from the US National Institutes of Health.

Cases such as Chang's illustrate that retracted papers need not doom a career, even a fledgling one. But navigating the process requires a willingness to deal with policies and distinctions that are not always clear-cut. "The greatest currency we have as scientists is respect — that colleagues respect us and our work," says Casadevall. "A retraction is survivable — if it's tackled honestly and transparently." ■

Virginia Gewin *is a freelance writer in Portland, Oregon.*

1. Lee, S.-W. *et al. Science* **326,** 850–853 (2009).
2. Han, S.-W. *et al. PLoS ONE* **6,** e29192 (2011).
3. Fang, F. C., Steen, R. G. & Casadevall, A. *Proc. Natl Acad. Sci. USA* **109,** 17028–17033 (2012).
4. Mirouse, V., Swick, L. L., Kazgan, N., St Johnston, D. & Brenman, J. E. *J. Cell Biol.* **177,** 387–392 (2007).
5. Mirouse, V., Christoforou, C. P., Fritsch, C., St Johnston, D. & Ray, R. P. *Dev. Cell* **16,** 83–92 (2009).
6. Haack, T., Bergstralh, D. T. & St Johnston, D. *Biol. Open* **2,** 1313–1320 (2013).
7. Swick, L. L., Kazgan, N., Onyenwoke, R. U. & Brenman, J. E. *Biol. Open* **2,** 1321–1323 (2013).
8. Lu, S. F., Jin, G. Z., Uzzi, B. & Jones, B. *Sci. Rep.* **3,** 3146 (2013).
9. Makarova, T. L. *et al. Nature* **440,** 707 (2006).
10. Jin, G. Z., Jones, B., Lu, S. F. & Uzzi, B. *The Reverse Matthew Effect: Catastrophe and Consequence in Scientific Teams* (National Bureau of Economic Research, 2013).

COLUMN

Strapped students

Graduate students face myriad financial challenges — some more obvious than others, says **Susie Crowe**.

Postgraduate education is a minefield of financial pitfalls — from tuition fees and loans to long stretches without contributing to savings. And yet many graduate students neglect to plan their finances accordingly.

The difference between starting a retirement plan in your early 20s versus your early 30s is in the hundreds of thousands of dollars. This realization hit me like a ton of bricks when I attended a seminar at my husband's workplace. At 28, I had long missed my opportunity to take full advantage of compound interest. Like many graduate students, I had spent my early 20s focusing on day-to-day financial challenges rather than on the big picture.

There are smaller, often overlooked financial traps. They include interest charges from living off credit cards; fees for late tuition payments; library fines; parking expenses; and unintentionally paying for your own research because you have lost receipts for fieldwork or laboratory expenses (see *Nature* **501,** 579–581; 2013). These seem trivial and short term compared with larger concerns, but they can mean the difference between comfortably paying your bills and barely scraping by.

KOYA979/SHUTTERSTOCK

Many expenses can be avoided or mitigated if you stay organized and remain on top of your finances. But in some cases, this is not so simple.

The money I am paid comes from so many sources it makes my head spin. Most students hope that income from teaching-assistant and research-assistant posts, external awards, internal scholarships, one-time entrance scholarships, tuition waivers and reimbursements will outweigh tuition and other fees, research costs, administrative charges, union dues and 'clawbacks' (funds that must be returned to a university administration owing to accounting errors). It can be tough for even the most financially savvy graduate student to keep a handle on things.

These days, my biggest worry is that my degree will drag on for longer than my current funding — a fate I have witnessed often enough to make every thesis setback seem like an ominous premonition. Sinking thousands of dollars into tuition for extra semesters is probably not a scenario envisioned by eager new students, but it is one of the more serious financial risks they can incur.

Given this bleak picture, what is a financially conscious student to do? The tired cliché of the impoverished graduate student is trotted out like a comforting inside joke, but it does not have to be the case. It is possible to make it through postgraduate education in decent financial shape.

Getting a head start in research was more important to my financial solvency than I initially appreciated. I got a job as a field assistant for an ornithology project early in my undergraduate studies, which gave me the skills to think up my own thesis project and led to a first-author publication in *Animal Behaviour*. This project gave my CV the boost it needed to get me a competitive national scholarship.

I try to be frugal and forward-thinking. It is important to be mindful of how tax laws apply to you — for example, in Canada, full-time students do not pay taxes on scholarships, but part-time students do. Choose a retirement plan early, and pay into it whatever you can. With compound interest, even the smallest efforts can yield big returns.

'Work hard and work smart' is a useful mantra. Efficiency buys time that can be spent on side jobs or business ventures to provide supplemental sources of income and a back-up plan if things go wrong. Students who stay focused and graduate quickly will avoid sinking money into extra tuition.

Be realistic. For prospective students put off by the financial difficulties that they may need to endure, I suggest thinking critically about whether a PhD will benefit you. Do the intellectual and career benefits seem worth the struggle? The hard truth is that if a prospective student has trouble defining what makes a PhD 'worth it' in the long run, then perhaps it is not. ■

Susie Crowe *is a doctoral candidate in biology at Carleton University in Ottawa, Canada.*

MARKET FORCES

Paradise revamped.

BY IAN STEWART

When Andrew Gordon was crossing the railway lines and failed to notice the approaching InterCity express, he didn't have time to think about death. But if there had been time, he would, as an atheist, have thought: "This is the end."

Clearly it wasn't.

He found himself standing in an eerie, luminous landscape. A narrow path wound its way through banked clouds to an elaborate gateway, where a figure in gleaming white robes waited. He wore a halo. Short stubby wings sprouted from his shoulders.

This may require a rethink. "Saint Peter?"

"Security pass, please, sir."

Andrew managed a shaky "Sorry, I don't have a pass."

"Ah. New soul in the Cloud." The man cocked his head, as if listening. "Gordon, Andrew Donald? 15 Wysteria Way, Dorridge?" Andrew nodded. "We don't always get notified, you see. Omniscience is all very well, but … too much information. Accident, was it?"

"No idea. Never saw it."

Another cock of the head. "Train. 10:43 from Wolverhampton. Running late. Hard luck."

Something here wasn't right. "You *are* Saint Peter? These are the pearly gates?"

The man shook his head. "You're thinking of the previous administration, son. We don't do sainthood any more."

"But — your halo. Wings."

"Retro look. This month's promotional initiative from Marketing. This is the Security Entrance, and I'm Pete the Security Guard. We don't let any old riff-raff in."

"Only true believers? I have a horrible feeling —"

"Belief? No, that's obsolete. What matters now is added value on your soul."

"I always thought souls were a mistaken reification of patterns of brain activity. You mean I really do have a soul?"

"It's more a case of your soul has got you." Seeing Andrew's baffled stare the guard added, "You're a *Christian* atheist, that's your problem."

"I'm not a Christian!"

"No, but the God you don't believe in is the Christian God."

Andrew bristled. "I don't believe in *any* gods." He looked sheepish. "Well, I didn't."

"Sure. But the *main* one you disbelieved in was the one your culture tried to *get* you to believe in. That coloured your expectations for the afterlife that you also don't believe in. Not

total nonsense, but … anyway, omnipotent or not, the Old Man adopted a new business model. He's now CEO, Chairman and CFO of Paradise Group. Holy Trinity, right? The afterlife is now a business, not a public service. The archangels have become the board of directors; archbishops and the Pope are relationship managers down in the Mundane. We had to let the cherubim and seraphim go, of course."

"What about Hell?"

"Hades Inc.? Our main competitor before Belle's new strategic vision." Seeing Andrew's puzzlement, he added: "Belle Z. Bubb, Hades' former Director of Human Resources, now CEO, Chair, CFO *and* Director of Inferno Technology as well."

"The Devil is *female*?"

"Belle's not exactly — look, there's no glass floor any more."

"You mean glass ceiling," Andrew said in reflex.

"No, I don't," said Pete, looking down through the clouds. "Where were we? Oh, yeah, souls. A soul isn't a *thing*, Andy. It's a spiritual instrument. An option on *you*, realized at death. Could be a call option, could be a put option."

"Sorry, I don't understand the jargon."

"Call option gives the right to buy at a set price; put option gives the right to insist that the other party buys."

"What caused the changes?"

"Lack of regulatory oversight, lavish bonuses, loss of stakeholder power, loss of employee power, third deadly sin …" He leaned closer to whisper in Andrew's ear.

"Actually, mate, I reckon the new corporate structure's worse than the old ways. Dodgy accounting, perverse incentives. Most trade now is in derivatives. Conduct Default Swaps, Innocence Rate Caps … not enough actual *souls* to justify the bonuses, you see. So the whole enterprise is built on sand, to quote the Chairman. A few years ago both companies nearly went belly-up because they'd accumulated a speculative bubble in CDOs."

"CDOs?"

"Collateralized Death Obligations. You take out an option on a portfolio of souls with bodies as collateral, package them together, trade the packages. Trouble was, no one knew what the things were worth. Thought it didn't matter: they could always realize the collateral assets. But when the market went belly-up and dead bodies were being dumped by the million, their spiritual value collapsed. Then Belle and the Old Man decided it was inefficient to compete, and went for horizontal integration. Merger in all but name."

"What about good and evil? They surely didn't merge those?"

"Course not. Under the previous management structure, our primary stakeholder base believed that whatever the Old Man said was good, *was* good. That gave authority to their concepts of morality and ethics and justice. Apparently. We still follow that principle. Anything that the directors of Paradise Group approve is good, by definition. Course, as part of the merger, same goes for Hades Inc.. It does save a lot of argument."

"But — what about people?"

"Not sure what you're driving at, son."

"What do ordinary people gain from all this … commercialization?"

Pete looked shocked. "*Gain* from it? Nothing. Why should they? *People* aren't on the corporate radar. They're commodities now, not customers."

"But Pete, *I'm* people. What will happen to me?"

The Security Guard sighed. "Well, Andy … now you're dead, your soul — the contract on *you* — matures. The option position unwinds, and …" He cocked his head for the third time, pursed his lips in satisfaction. "Got that, Gabe. OK, Andy: the good news is, your soul is currently part of Paradise Group's portfolio."

Andrew almost collapsed with relief. Then he realized he might not have the complete picture. "And the bad news?"

"It's a put option." ■

Ian Stewart, *emeritus professor at the University of Warwick, writes popular science books and science fiction.*

JACEY